SPARROW Surface Water Quality Workshop

October 29-31, 2002

Reston, Virginia

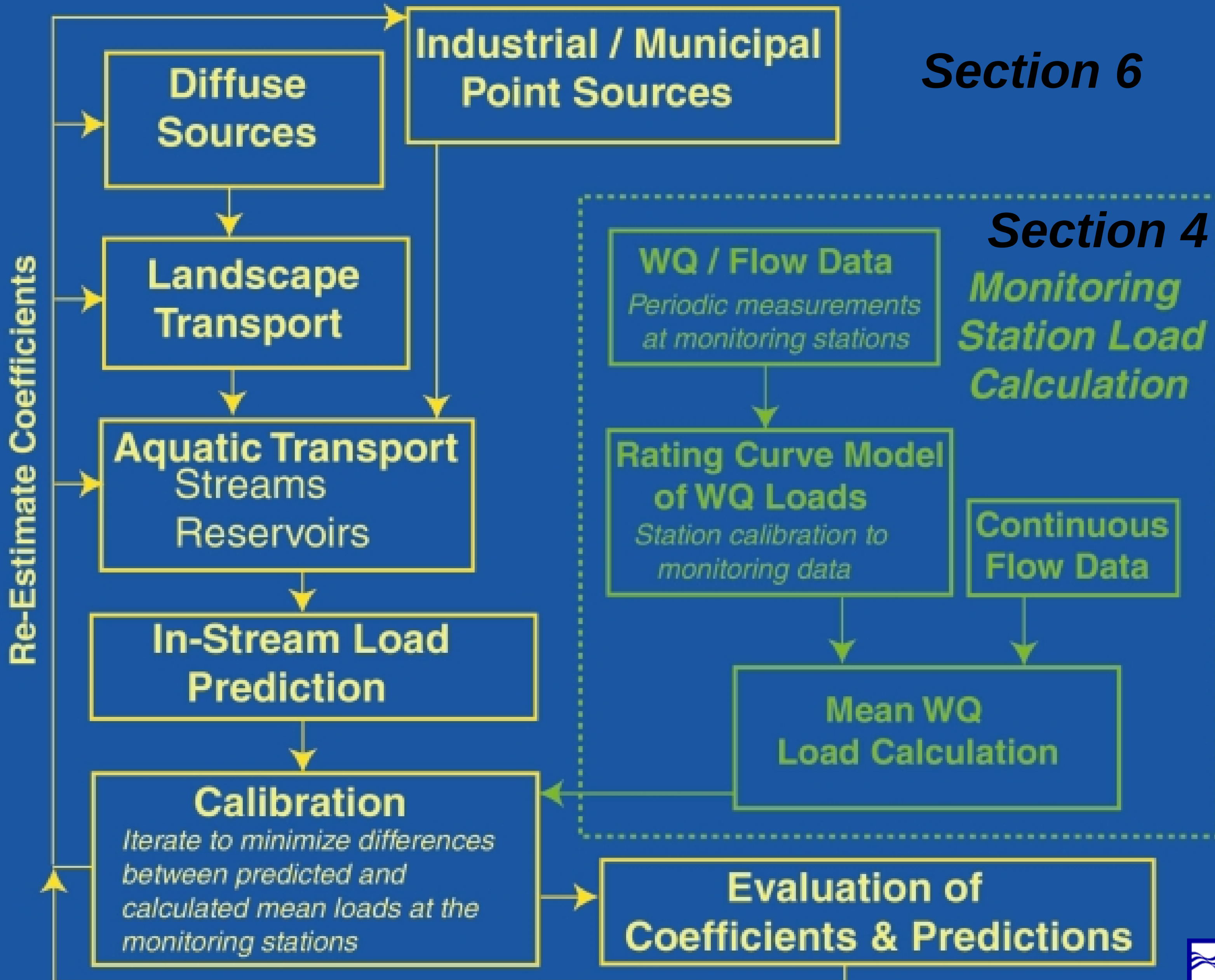
Section 6. SPARROW Model Calibration

USGS
science for a changing world

Section 6. SPARROW Model Calibration Topics Considered

- The SPARROW model equation
- Specification of the model:
  sources, land-to-water and aquatic transport
- Nonlinear estimation of parameters
- Physical interpretation of parameters
- Evaluating the model error
- Model selection criteria
- SPARROW calibration/prediction software (existing)
  Fecal coliform example
- On-going enhancements to the code

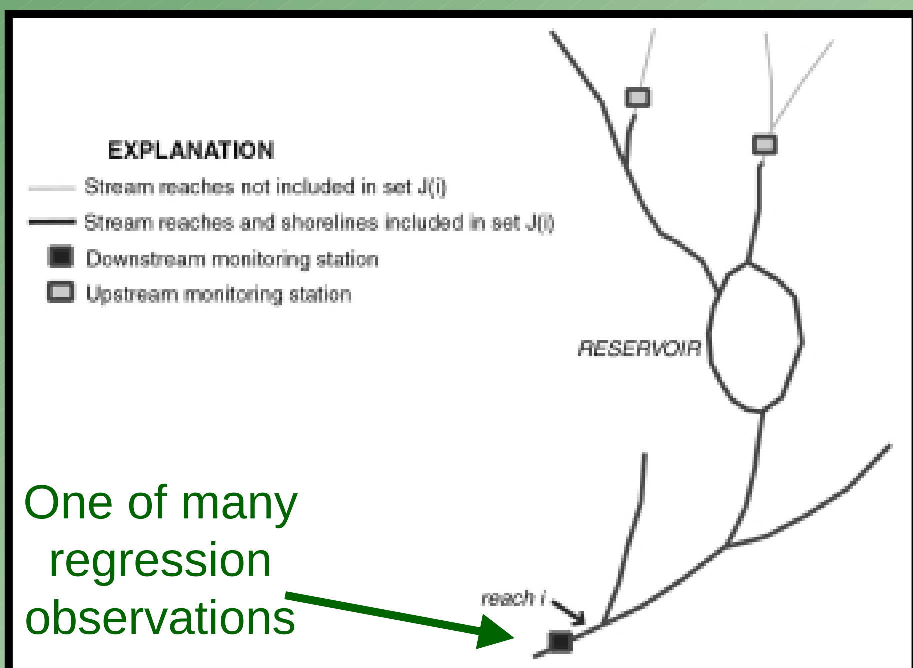
SPARROW Model Structure



USGS
science for a changing world

SPARROW Model Equation

$$\underbrace{\text{Load}_i}_{\text{Stream Load at reach } i} = \left\{ \sum_{j \in J(i)} \left[\sum_{n=1}^{N} \underbrace{S_{n,j}}_{\text{Sources}} \underbrace{\beta_n \exp(-\alpha' Z_j)}_{\text{Land-to-water transport}} \right] \underbrace{\exp(-\delta' T_{i,j})}_{\text{Aquatic transport}} \right\} \underbrace{\exp(\epsilon_i)}_{\text{Error}}$$



n = number of sources
j = reaches in the set J(I)

SPARROW Model Equation

<u>Mass-Balance Equation</u>

$$\underset{\text{Stream Load}}{\mathrm{Load}_i} = \left\{ \sum_{j \in J(i)} \left[\sum_{n=1}^{N} \underset{\text{Sources}}{S_{n,j}}\, \beta_n \underbrace{\exp(-\alpha' Z_j)}_{\text{Land-to-water transport}} \right] \underbrace{\exp(-\delta' T_{i,j})}_{\text{Aquatic transport}} \right\} \underbrace{\exp(\epsilon_i)}_{\text{Error}}$$

<u>Model structure is nonlinear:</u>

- Additive sources (preserves mass balance)
- Multiplicative error (account for scale dependency of error)
- Exponential delivery terms

Model Equation

How does the model structure differ from that of a conventional log-linear watershed regression model?

Nonlinear Model: $Y = (X_1B_1+X_2B_2)\, e^{\varepsilon}$

Log transform: $\ln(Y) = \ln(X_1B_1+X_2B_2) + \varepsilon$

- Additive sources
- Mass balance
- Multiplicative error

Log-Linear Model: $Y = (X_1{}^{B}{}_1\, X_2{}^{B}{}_2)\, e^{\varepsilon}$

Log transform: $\ln(Y) = B_1\ln(X_1)+B_1\ln(X_1)+\varepsilon$

- Multiplicative sources
- No mass balance
- Multiplicative error

USGS

science for a changing world

SPARROW Model Sources

$$\mathrm{Load}_i = \left\{ \sum_{j \in J(i)} \left[\sum_{n=1}^{N} S_{n,j}\, \beta_n \underbrace{\exp(-\alpha' Z_j)} \right] \underbrace{\exp(-\delta' T_{i,j})} \right\} \underbrace{\exp(\epsilon_i)}$$

Stream Load

Sources

Land-to-water transport

Aquatic transport

Error

SPARROW Model Sources

- Selection of sources determined by:
  - Research literature
  - Your expertise – knowledge of the watershed
  - Data availability
- Diffuse Sources:
  - Extensive (land-use)
  - Intensive (mass based)
  - Mixed model (extensive and intensive)
- Point sources:
  - Contaminant mass (expect coefficient of 1.0 if response variable in same units and model properly specified)
  - Surrogates: sewered population; BOD; Flow
- Geographic dummy variables (e.g., unspecified sources)

SPARROW Landscape Transport

$$\text{Load}_i = \left\{ \sum_{j \in J(i)} \left[\sum_{n=1}^{N} S_{n,j} \beta_n \underbrace{\exp(-\alpha' Z_j)}_{\text{Land-to-water transport}} \right] \underbrace{\exp(-\delta' T_{i,j})}_{\text{Aquatic transport}} \right\} \underbrace{\exp(\epsilon_i)}_{\text{Error}}$$

Stream Load

Sources

Land-to-water transport

Aquatic transport

Error

SPARROW Model Landscape Transport

Selection of landscape variables:

- Variables should relate closely to landscape processes (e.g., runoff and drainage area integrate terrestrial and aquatic processing and water transport—not recommended)
- Water balance inputs and landscape-related loss components (precipitation; evapotranspiration)
- Soil properties (e.g., organic content, permeability, moisture content)
- Water flow paths (e.g., TOPMODEL overland flow; DEM overland routing)
- Management activities (e.g., tile drainage, conservation tillage practices; BMPs—stream riparian properties)
- Land use (e.g., wetlands—is it a source or sink?; measures of impervious surface or urbanization may confound source estimation)

SPARROW Model Landscape Transport

Landscape decay functions:

- Exponential function with imbedded negative sign: constrains values between zero and one:
  - --All coefficients reported with positive sign
  - --Negatively related variables:

$$e^{(-\beta_1 X_1 - \beta_2 X_2)}$$

(e.g., soil permeability)

  - --Positively-related variables enter as reciprocal:

$$e^{(-\beta_1 1/X_1 - \beta_2 1/X_2)}$$

(e.g., drainage density)

- Exponential function with log transformed variables – unconstrained:

$$e^{(\beta_1 \log(X_1) + \beta_2 \log(X_2))}$$

  - --Coefficients reported with actual sign

USGS
science for a changing world

SPARROW Model Landscape Transport

Landscape-source variable interactions:

• Values of diffuse-source coefficients not independent of delivery variables (interaction not separable in model)

$$\text{Load}_i = \left\{ \sum_{j \in J(i)} \left[\sum_{n=1}^{N} S_{n,j} \underbrace{\beta_n \exp(-\alpha' Z_j)} \right] \underbrace{\exp(-\delta' T_{i,j})} \right\} \underbrace{\exp(\epsilon_i)}$$

• Standardize delivery variables (deviations from mean) to create more interpretable source coefficients expressed in relation to mean delivery—yields reflecting delivery to streams best metric to reflect geographic variations

• Commonly assume that diffuse sources subject to same landscape and aquatic decay (however, code can accommodate separate source-delivery interactions)

SPARROW Aquatic Transport

$$\mathrm{Load}_i = \left\{ \sum_{j \in J(i)} \left[\sum_{n=1}^{N} S_{n,j} \underbrace{\beta_n \exp(-\alpha' Z_j)}_{} \right] \underbrace{\exp(-\delta' T_{i,j})}_{} \right\} \underbrace{\exp(\epsilon_i)}_{}$$

Stream Load

Sources

Land-to-water transport

Aquatic Transport (streams and reservoirs)

Error

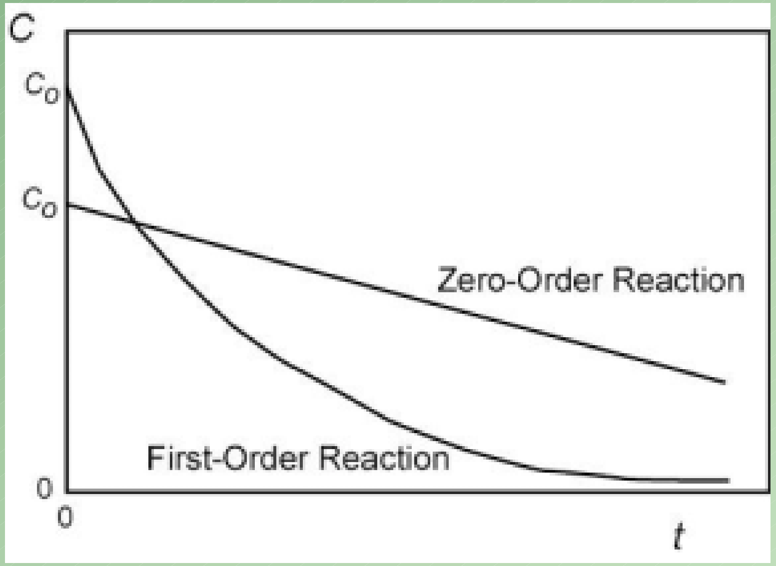
USGS
science for a changing world

Stream Transport

Chemical Reaction Kinetics

–The reaction rate is proportional to the concentration of the reactants (law of mass action):

$$dc / dt = -kc^n$$



k = loss rate (T^{-1})

c = concentration of a single reactant

n = reaction order

–Integrated equation for the first-order model (exponential depletion):

$$c = c_0 e^{-kt}$$

USGS science for a changing world

SPARROW Stream Nutrient Transport

$$\text{Flux} = \text{Flux}_0\, e^{-kt}$$

Loss rate (per unit water travel time) integrates multiple processes:

$$k = k_D + k_U + k_S$$

where,

k_D = denitrification rate (N only) ~ f[benthic areal rate(+), depth(-), concentration(+/-), temperature(+), organic matter(+), flow(-)]

k_U = biological uptake rate ~ f[algal density (+), light(+), depth(-), concentration(+), temperature(+)]

k_S = settling loss rate ~ f[particle settling velocity (+), depth(-), particulate bound fraction (+)]   (particulate burial?)

t = water time of travel

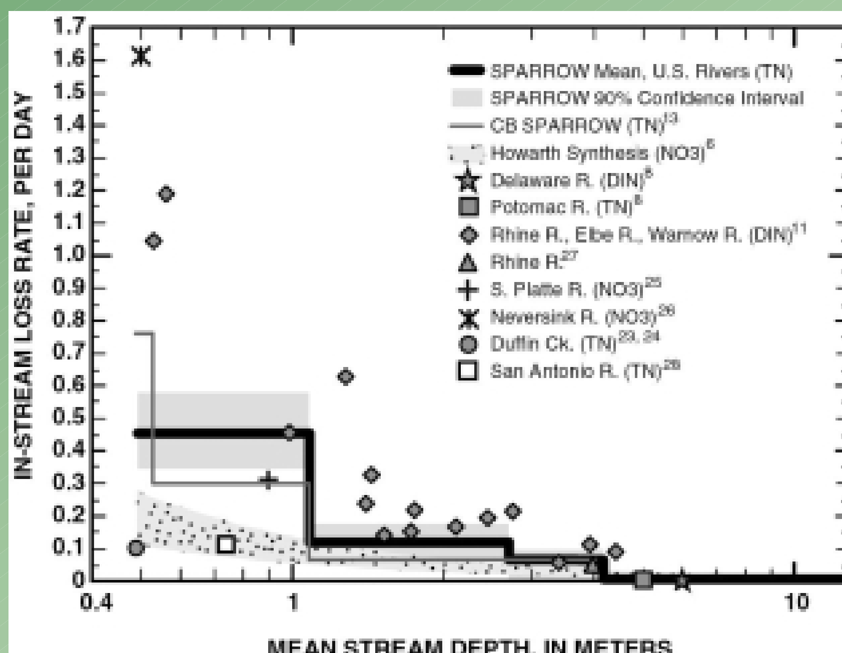
Related Questions: time scale of steady state processes (multiple species model) and spatial scaling factors

SPARROW Stream Nutrient Transport

<u>Depth (streamflow) theoretically important spatial scaling factor affecting nutrient loss rate:</u>

$$\text{Flux} = \text{Flux}_0 \, e^{-k_n t}$$

where, *n* = flow class defined according to mean streamflow

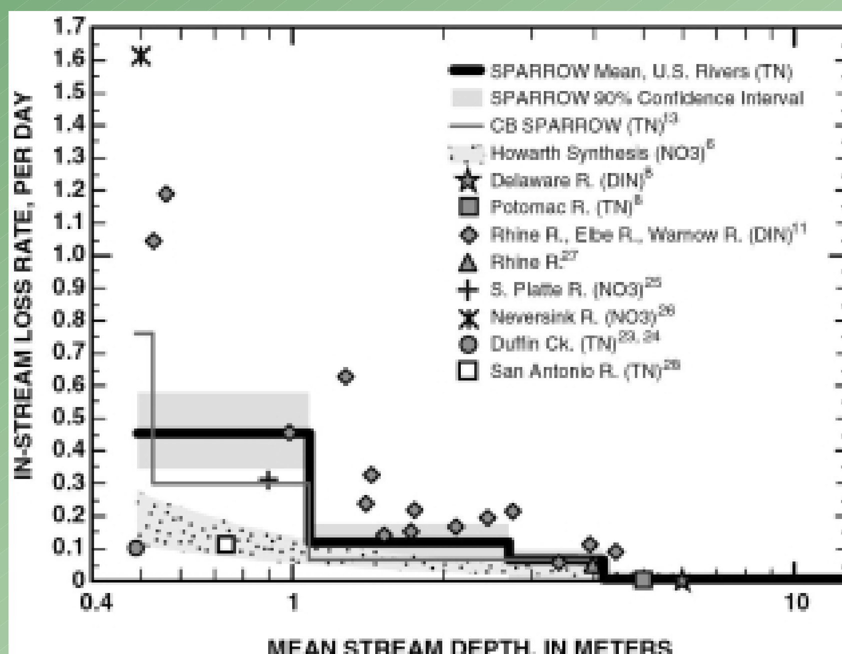


SPARROW Estimates of Total Nitrogen Loss vs. Channel Depth

SPARROW Stream Nutrient Transport

Approach for defining flow classes:

- –Use a few integer factor and order of magnitude separations
- –Start simple with 2 to 3 flow classes
- –Evaluate final selections with continuous flow functions



SPARROW Estimates of Total Nitrogen Loss vs. Channel Depth

SPARROW Stream Nutrient Transport

Per unit time (day^{-1}) rate: $Flux = Flux_0 \, e^{-k_n t}$

Per unit channel length (km^{-1}) rate: $Flux = Flux_0 \, e^{-k'_n L}$

Approaches to obtaining day^{-1} rate:

- Post-conversion: km^{-1} to day^{-1} rate using available estimates of mean stream velocity ($k = k' * V$) for comparison with literature values
- Pre-conversion: Stream morphological / hydrology studies (e.g., Leopold and Maddock, 1953; Jobson, 1996; Jowett, 1998) can be used to estimate time-of-travel for individual reaches based on streamflow

Obtaining estimates of channel depth:

e.g., Leopold and Maddock, 1953; Depth = $0.2612 \, Q^{0.3966}$

Nutrient Transport in Reservoirs/Lakes

Empirical mass-balance models:

- Vollenweider, 1969

  Phosphorus: Reckhow and Chapra, 1983

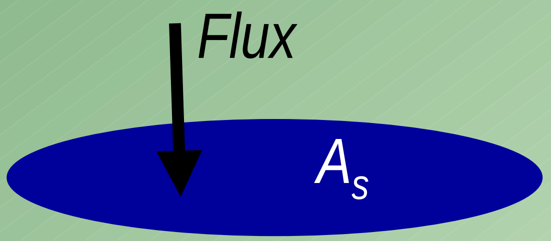
  Nitrogen: Kelly et al. 1987; Molot & Dillon, 1993

- Steady state, well-mixed conditions
- Retention ~ f(depth, residence time, volume, areal water load, and apparent settling velocity—mass transfer coefficient)

USGS
science for a changing world

Lake Retention of Nutrients

Depth-Independent vs. Depth-Dependent Approaches



Sediment-Water Interface

Constant Settling Velocity: Settling = $\nu A_s c$

ν = apparent settling velocity (units: L T^{-1})
A_s = lake surface area
c = concentration

First-order reaction (depth-dependent settling velocity):

Reaction = k_s V c

$k_s = \nu$ / depth; first-order settling rate (units=T^{-1})
V = volume (units=L^3=depth*surface area)

Which is better?

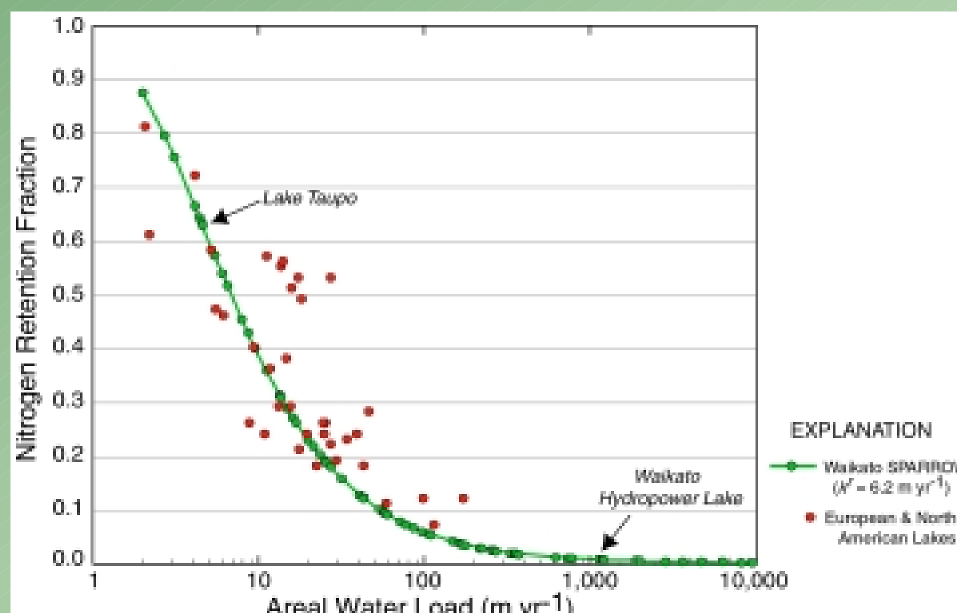
- Equivalent mathematically
- Former more specific to process of flux across sediment-water interface
- Little empirical evidence in lake literature of detectable differences
- Availability of data may determine (volume vs. surface area)

Nutrient Transport in Reservoirs/Lakes

<u>SPARROW estimated loss:</u>

- *Settling velocity coef.* (constant in all water bodies)
- *Areal water load* (ratio of mean outflow to surface area)

SPARROW Estimates of Total Nitrogen Loss
(75 Reservoirs of the Waikato River Basin, New Zealand)



Lake Retention of Nutrients
Current SPARROW Equation

From Mass-Balance Expression
(Depth Independent)
(e.g.,Reckhow and Chapra, 1983)

$$R = \nu / (\nu + q_s)$$

R = retention coefficient

ν = apparent settling velocity ($L\ T^{-1}$)

q_s = areal water load (outflow/surface area; $L\ T^{-1}$)

USGS
science for a changing world

Lake Retention of Nutrients
Previous SPARROW Equation

Empirical Approximation to Mass-Balance Expression

(Alexander et al. *Wat. Resour. Res.*, *in press*)

$$R = 1 - exp(-\nu/q_s)$$

R = retention coefficient

ν = apparent settling velocity ($L\ T^{-1}$)

q_s = areal water load ($L\ T^{-1}$)

USGS science for a changing world

Lake Retention of Nutrients
Original SPARROW Equation

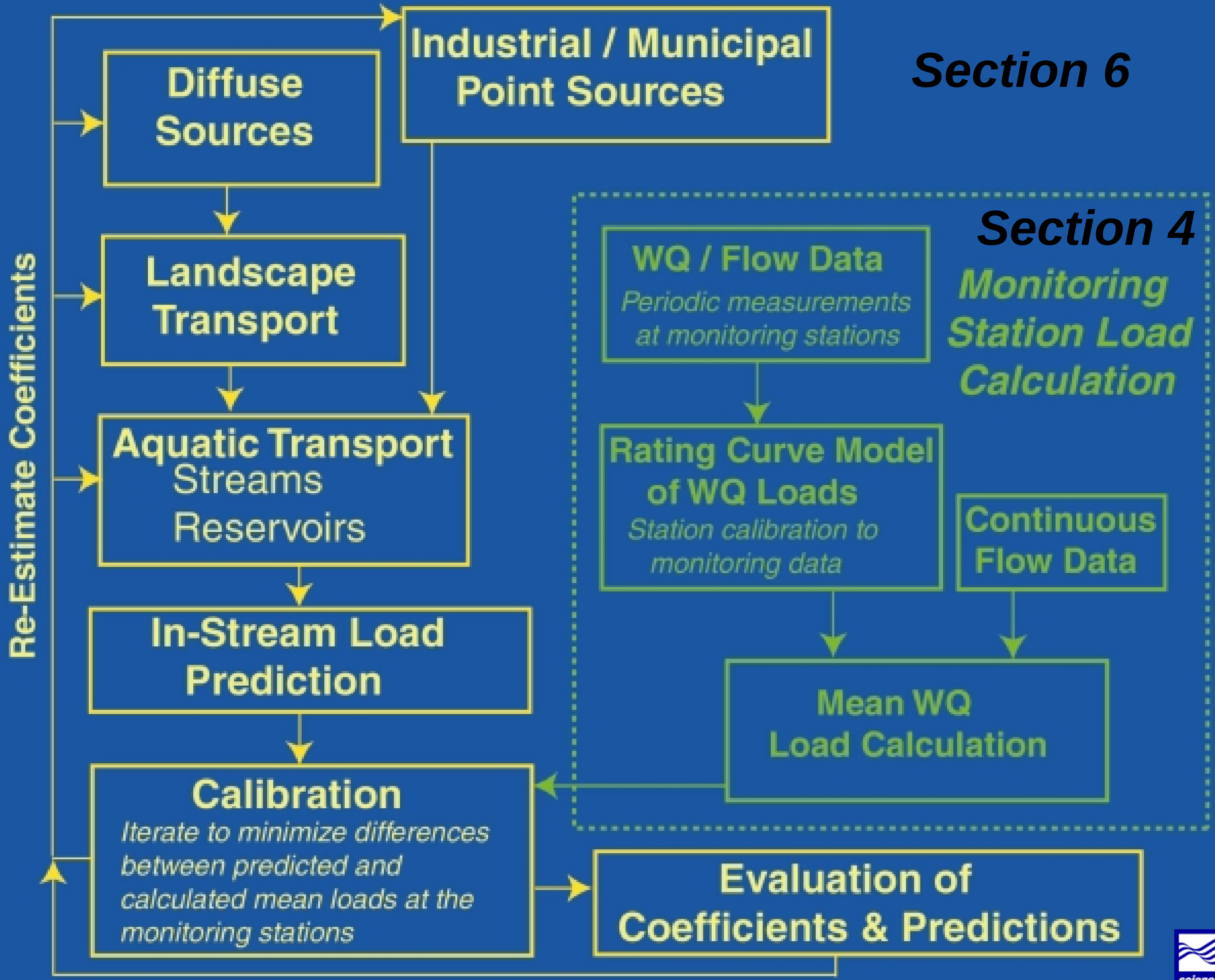
(Smith et al. *Wat. Resour. Res.*, *1997*)

$$R = 1 - exp(-kT)$$

R = retention coefficient
k = 1st-order decay rate
T = channel water travel time in
reservoir reach

USGS
science for a changing world

SPARROW Model Structure



USGS
science for a changing world

SPARROW Model Equation
Nonlinear Estimation of Parameters

$$\text{Load}_i = \left\{ \sum_{j \in J(i)} \left[\sum_{n=1}^{N} S_{n,j} \underbrace{\beta_n \exp(-\alpha' Z_j)} \right] \underbrace{\exp(-\delta' T_{i,j})} \right\} \underbrace{\exp(\epsilon_i)}$$

Stream Load

Source parameters

Land-to-water parameters

Aquatic parameters

Error

<u>Model structure is nonlinear:</u>

- Additive sources (preserves mass balance)
- Multiplicative error (account for scale dependency of error)
- Delivery terms

Nonlinear Parameter Estimation

Nonlinear Regression can be viewed as an extension of linear regression analysis.

Uses Gauss-Newton optimization method (Levenberg-Marquardt conditioning parameters)—an iterative form of standard linear regression

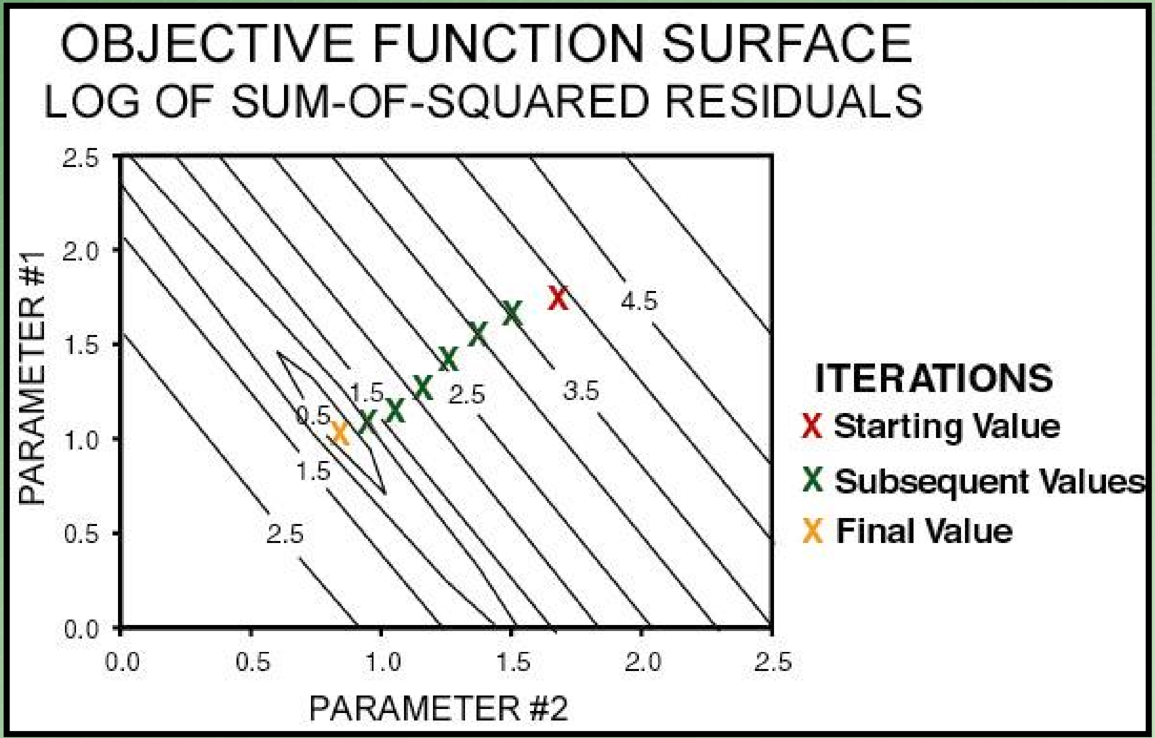
- User selects starting values for the parameters.
- Method iteratively applies linear approximations of the nonlinear model in the vicinity of the initial and subsequent parameter values until the model converges.

- Model converges when changes in the parameter estimates are less than a preset threshold. The parameters minimize the objective function.
- Objective function is a measure of the fit between predicted and observed values (e.g., sum-of-squared residuals)

USGS
science for a changing world

Nonlinear Parameter Estimation

- The optimization routine uses a linear approximation of the nonlinear objective function.
- Parameter values are changed iteratively to locate the minimum value of the objective function.



Nonlinear Parameter Estimation

How do you pick starting values?

- Literature values
- Other SPARROW models or local models
- Guestimate remaining values if lack information

Model convergence:

- If model well conditioned, only exceedingly large differences between starting and final values will be problematic
- Should test the convergence of final model for stability (selection of global rather than local minima) by changing starting values by a large amount
- SPARROW convergence problems often related to data errors; problems can occur when attempt to estimate too many parameters for certain functions (e.g., reservoir decay)

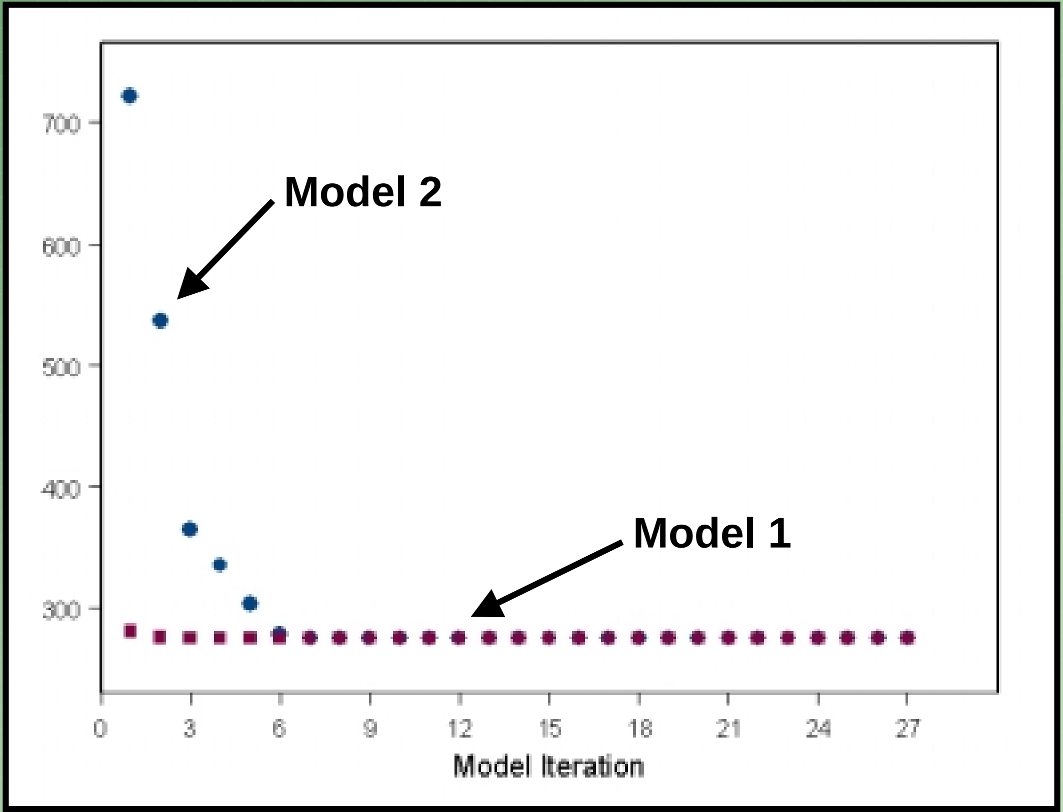
Nonlinear Parameter Estimation

Model Iterations of Fecal Coliform Model

Model 1: Starting values near final estimates (33 iterations)

Model 2: Starting values changed by 2 orders magnitude (aquatic decays by 1 order magnitude) – (27 iterations)

Value of Objective Function



Final parameter estimates identical for two models

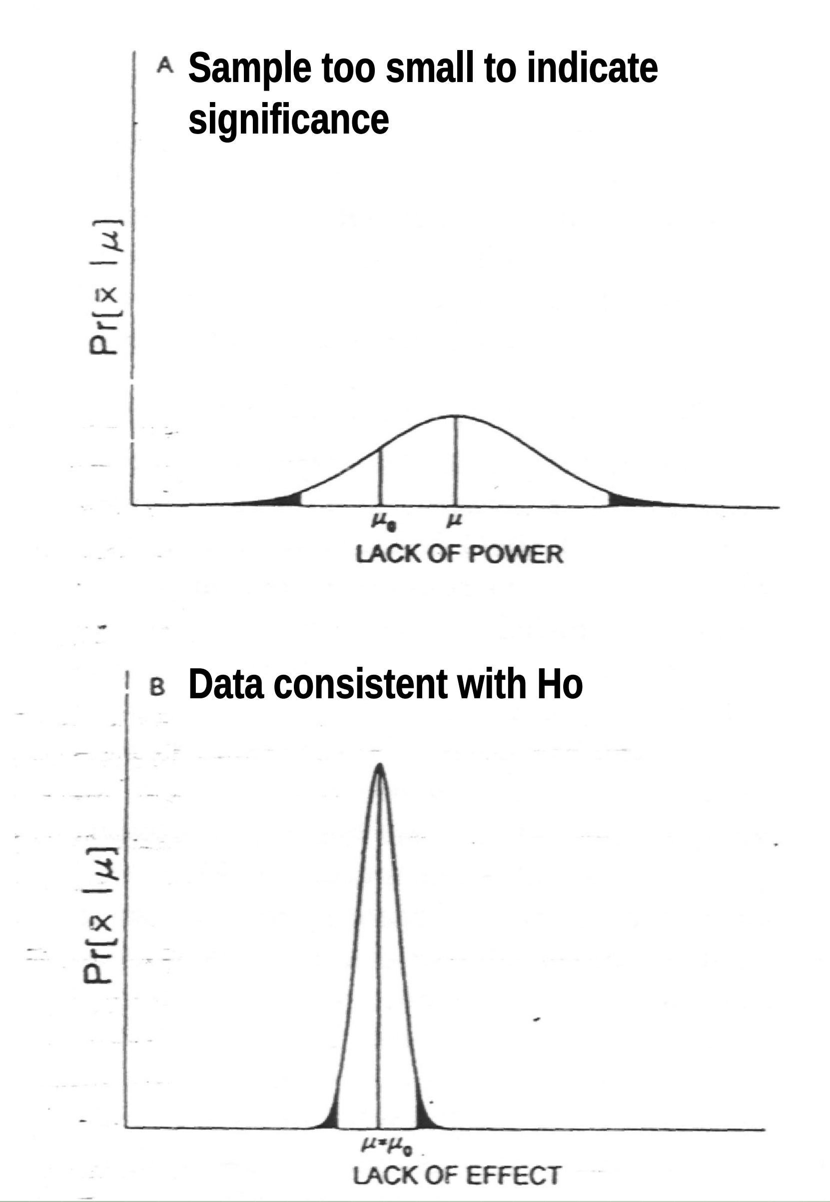
Nonlinear Parameter Estimation

<u>Check if converged model reasonable:</u>

- Do the parameter estimates have the correct sign?
- Are the parameters and standard errors statistically significant?
  - t-statistics: $t = b_j / se(b_j)$
  - tests of parameter significance:
    - t values approximate (only asymptotically valid)
    - strict adherence to α level not recommended
    - insignificant parameters: “lack of effect” vs. “lack of power”

USGS

science for a changing world



Lack of Power
vs.
Lack of Fit

Possible Interpretations of a Test That Fails to Reject Ho
(e.g., p>0.05)

From: Johnson, 1999, J Wildl. Manage., v. 63

USGS
science for a changing world

Nonlinear Parameter Estimation

Check if converged model reasonable:

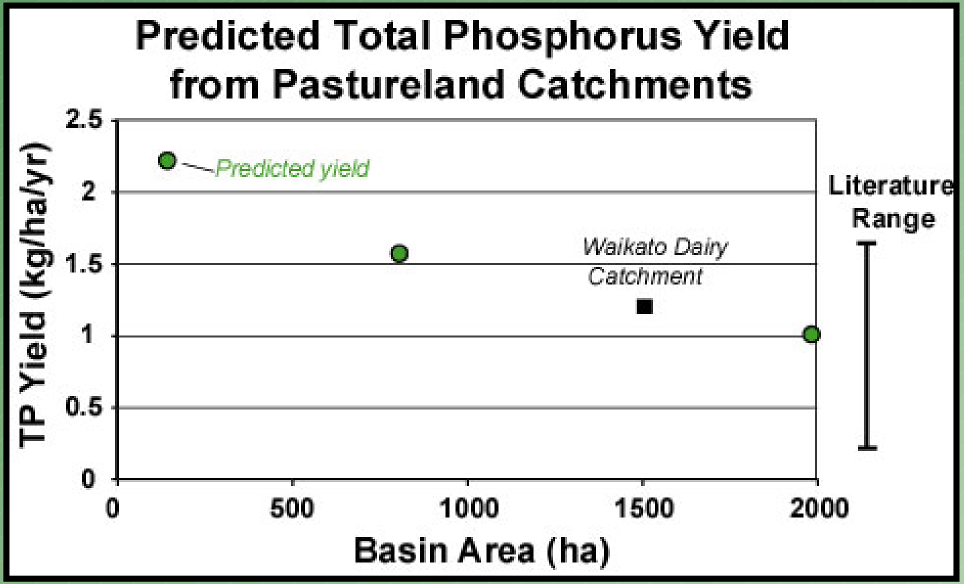
- Do the parameter estimates have the correct sign?
- Are the parameters and standard errors statistically significant?
    - t-statistics: $t = b_j / se(b_j)$
    - tests of parameter significance:
        - t values approximate (only asymptotically valid)
        - strict adherence to α level not recommended
        - insignificant parameters: “lack of effect” vs. “lack of power”
- Do the parameters have physical significance?

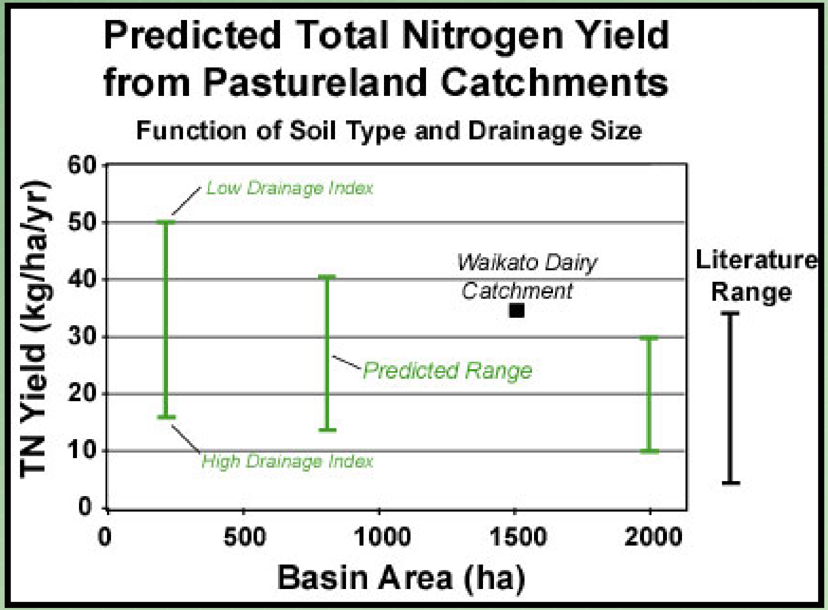
    Literature comparisons: catchment yields by land use, per capita waste loads, point-source coef., in-stream decay, reservoir settling rates

USGS
science for a changing world

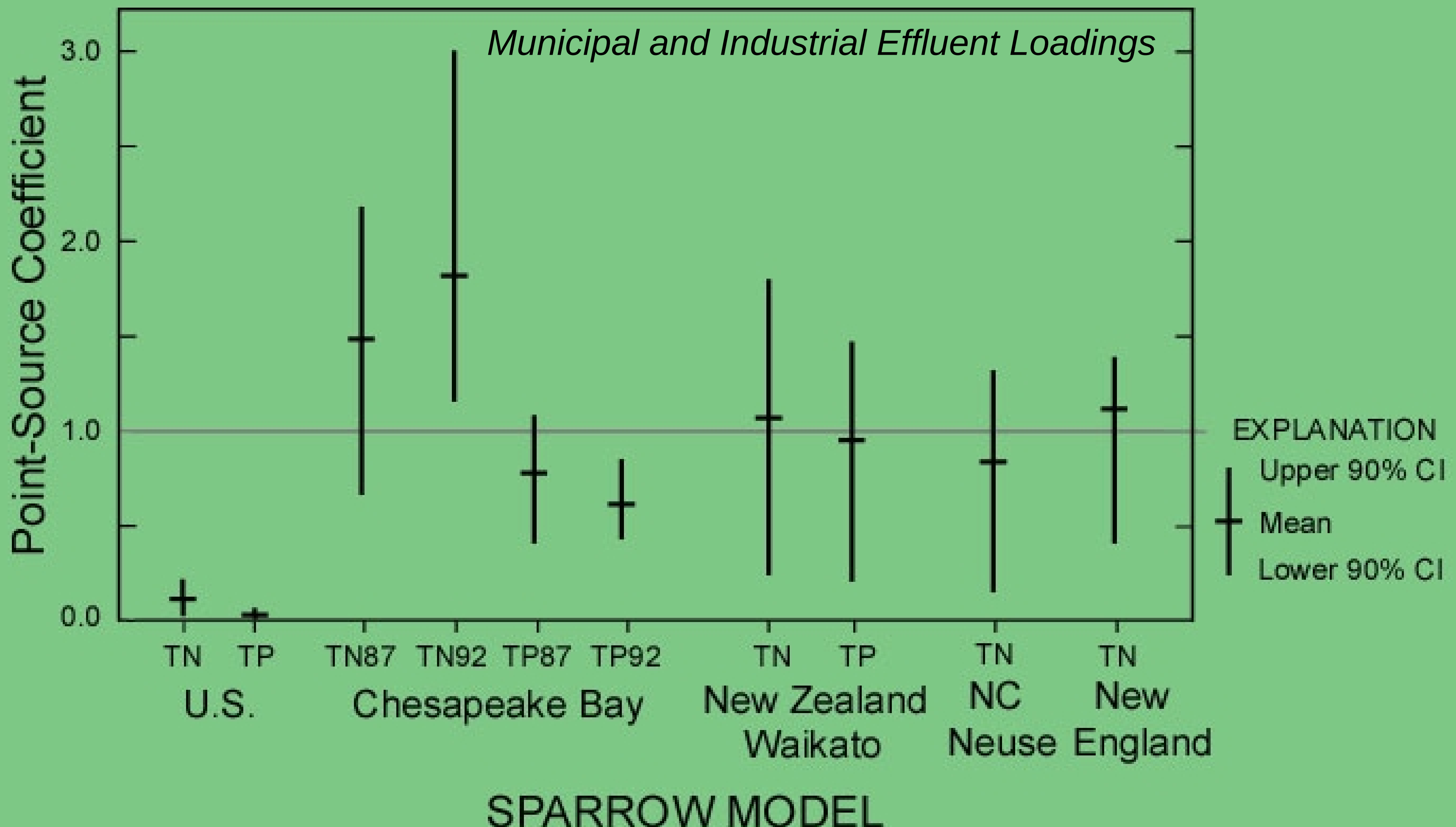
Verification of Estimated Diffuse Source Coefficients

New Zealand SPARROW

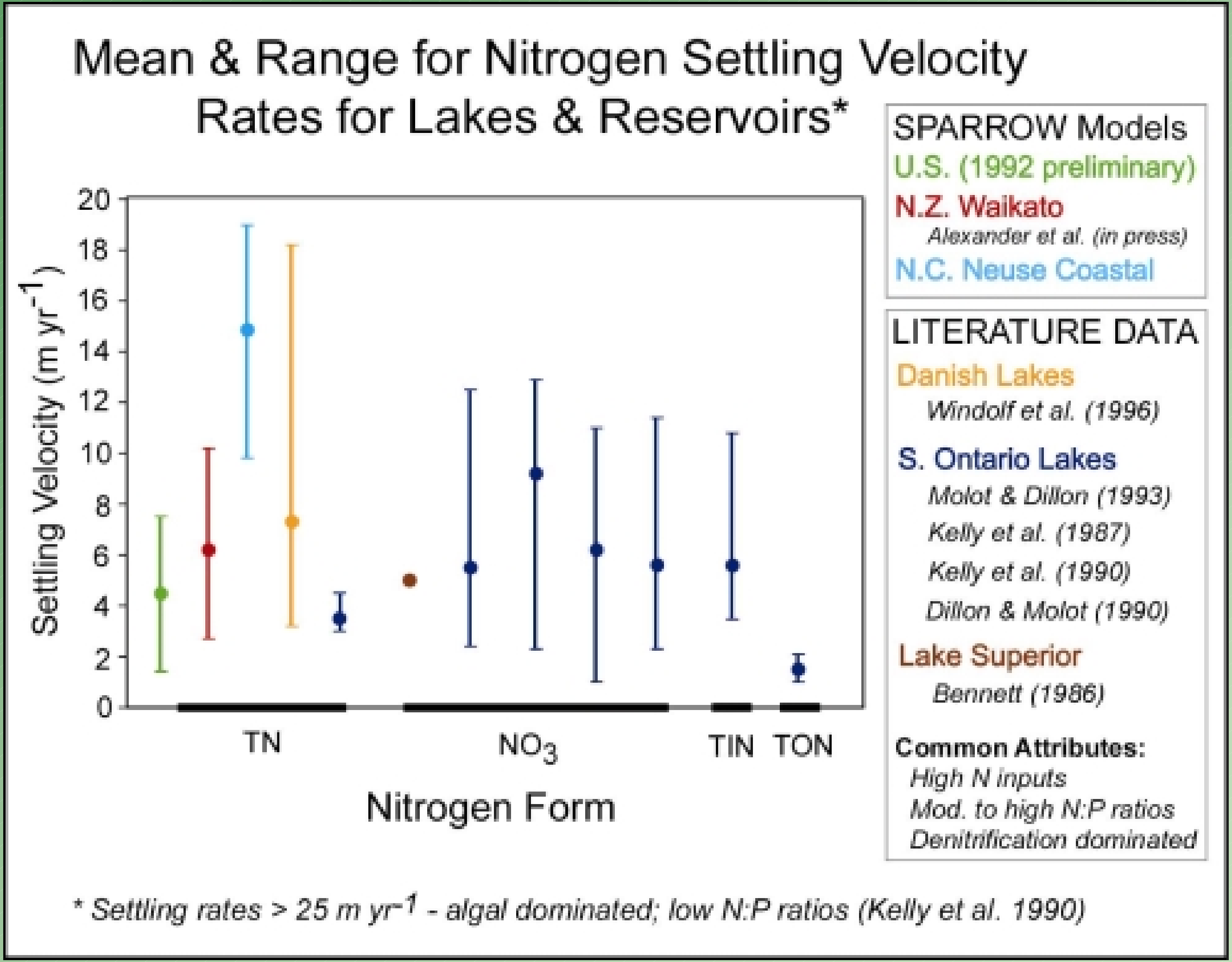




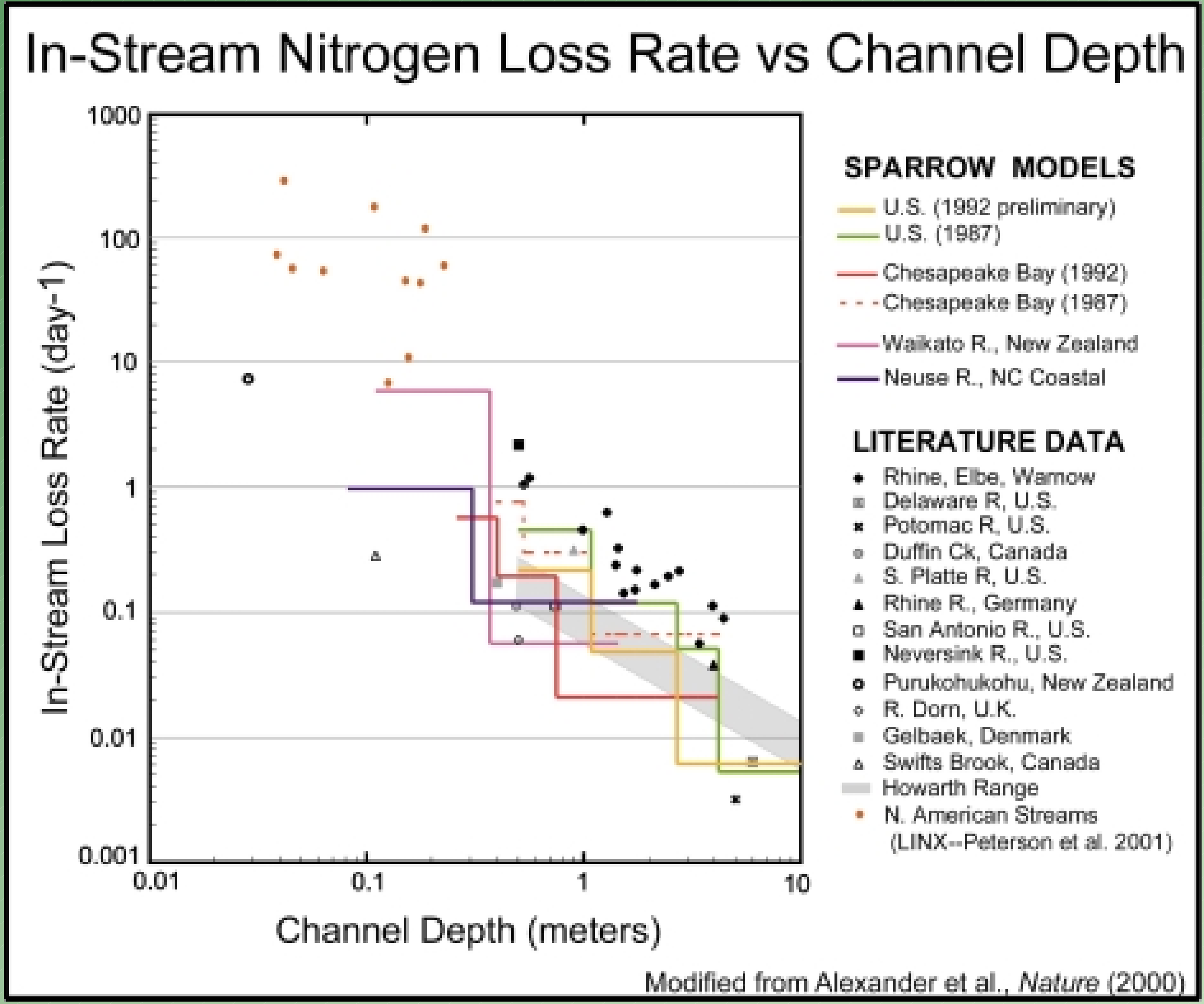
NATIONAL & REGIONAL SPARROW POINT-SOURCE COEFFICIENTS



SPARROW Estimates of Nitrogen Loss in Reservoirs



National and Regional SPARROW Models



Nonlinear Parameter Estimation

<u>Check if converged model reasonable:</u>

- Do the parameter estimates have the correct sign?
- Are the parameters and standard errors statistically significant?
    - t-statistics: $t = b_j / se(b_j)$
    - tests of parameter significance:
        - t values approximate (only asymptotically valid)
        - strict adherence to α level not recommended
        - insignificant parameters: "lack of effect" vs. "lack of power"
- Do the parameters have physical significance?

  Literature comparisons: catchment yields by land use, per capita waste loads, point-source coef., in-stream decay, reservoir settling rates
- Are the parameters correlated?

Parameter Correlation: Multicollinearity

What's the problem?

- Coefficient signs unreasonable
- Two variables describing the same process have different signs and insignificant coefficients

Metrics to detect it:

- Parameter correlation matrix – very high correlations (>0.95)
- Variance Inflation Factors (VIFs) – measure of correlation among all explanatory variables (>10 a problem)

How do you fix it?

- Center data – fix for polynomial variables
- Simplify model
- Remove parameter
- Combine data (equating two parameters)

USGS
science for a changing world

Nonlinear Parameter Estimation

Model convergence can also be assisted by:

• Log transformations of data and parameters

May be needed to satisfy residual assumptions (linearity, constant variance, normally distributed)

• Scaling of data – better conditioning of the derivative matrix if reduce large differences in coefficient values:

$$Y = b0 \exp(-b1\ X)$$

If expect b0=100 and b1=0.001, force coefficients to be approx 1.0:

$$Y = 100\ b0 \exp(-b1\ X/1000)$$

SPARROW Error Term

$$\text{Load}_i = \left\{ \sum_{j \in J(i)} \left[\sum_{n=1}^{N} S_{n,j} \beta_n \underbrace{\exp(-\alpha' Z_j)} \right] \underbrace{\exp(-\delta' T_{i,j})} \right\} \underbrace{\exp(\epsilon_i)}$$

Stream Load

Sources

Land-to-water transport

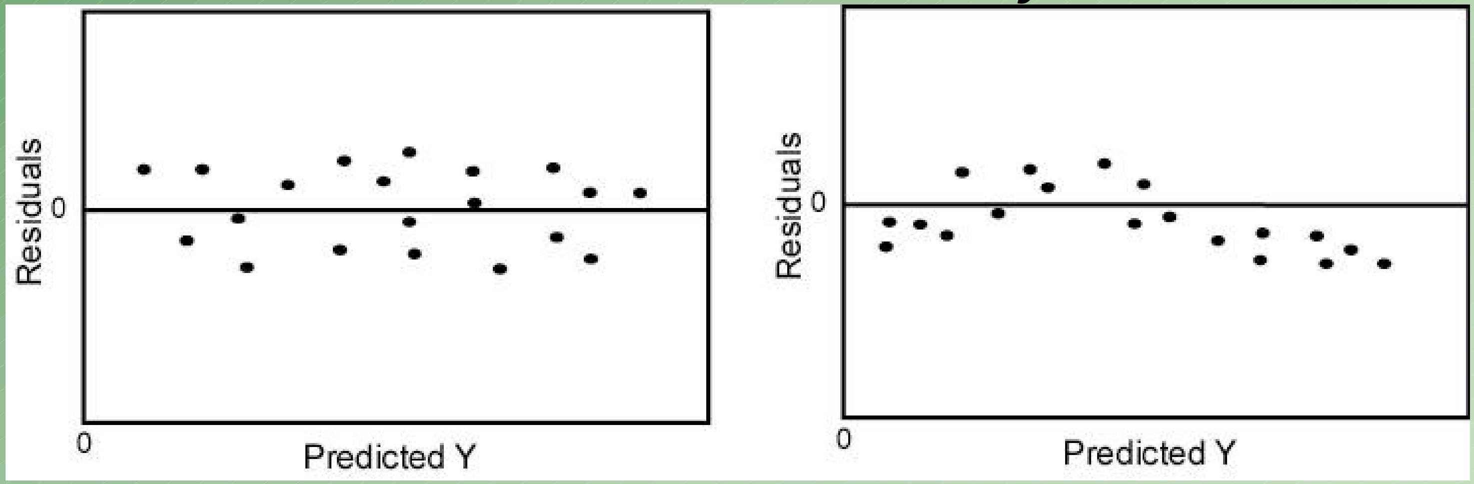
Aquatic Transport (streams and reservoirs)

Error

SPARROW Model Calibration Residual Analysis

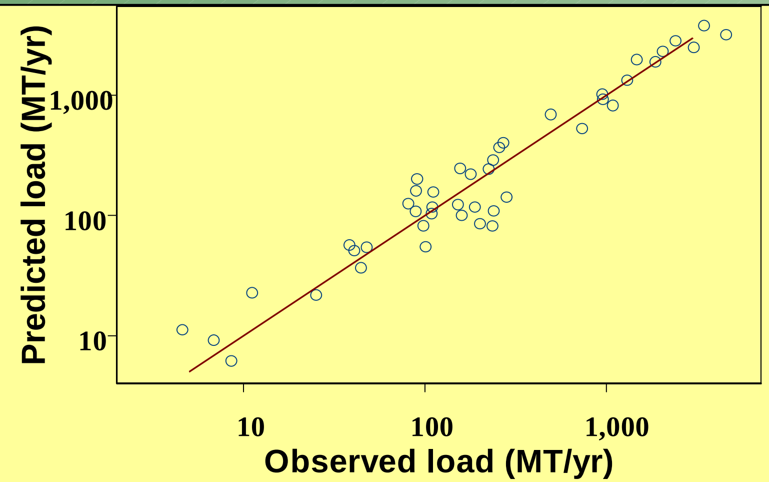
• Residuals should vary randomly – no evidence of systematic patterns that may indicate correlation or explanatory variables missing from the model

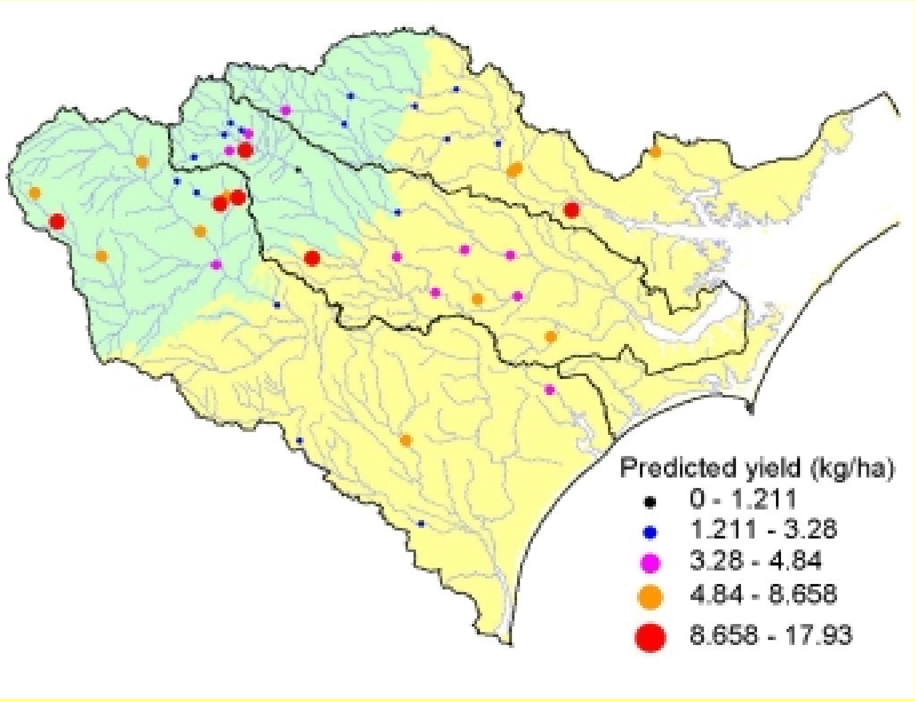
Good | *Not good, but the glass may be half full*



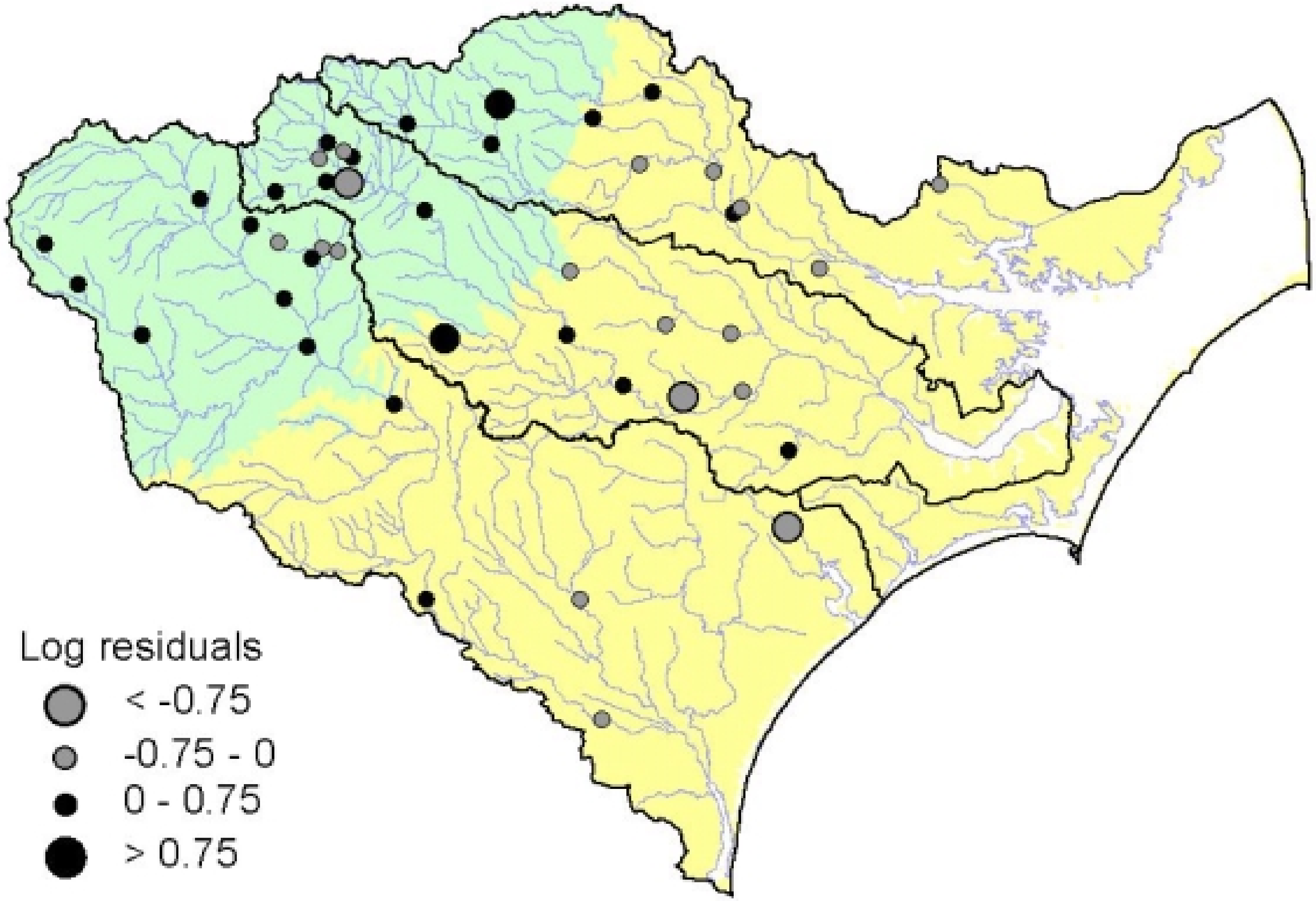
NC SPARROW land cover-based regression model

	Parameter	p-value
R2	**0.93**	
MSE	**0.22**	
TN sources		
Point sources (MT/1992)	0.85	0.006
Agr. Land area (MT/km^2)	5.9	0.09
Non-agr land area (MT/km^2)	1.79	0.08
Land delivery variable		
Soil hydrologic group	4.13	0.001
Aquatic loss		
Small stream (km^{-1})	0.08	0.02
Large stream (km $^{-1}$)	0.002	0.35
Reservoir (m/yr)	16.4	0.008





Spatial distribution of model residuals

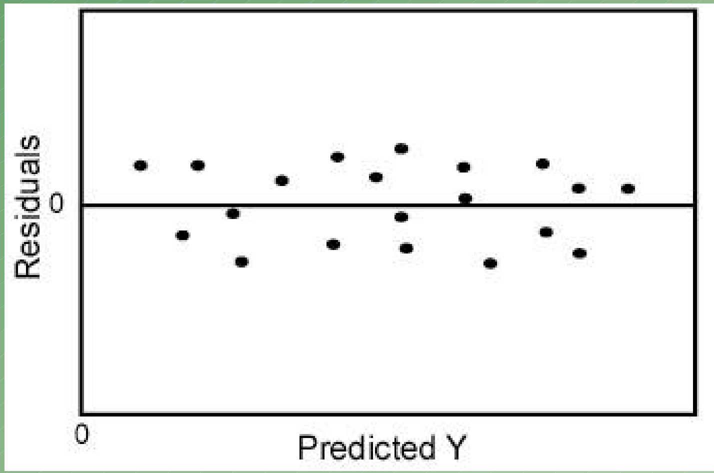


SPARROW Model Calibration
Residual Analysis

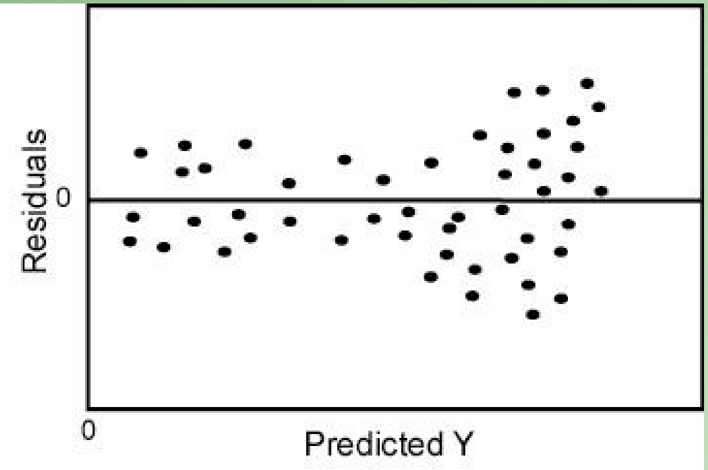
<u>Assumptions affecting accuracy of parameter standard errors</u>:

• Variance of the residuals should be constant (homoscedastic)

Good



Not good

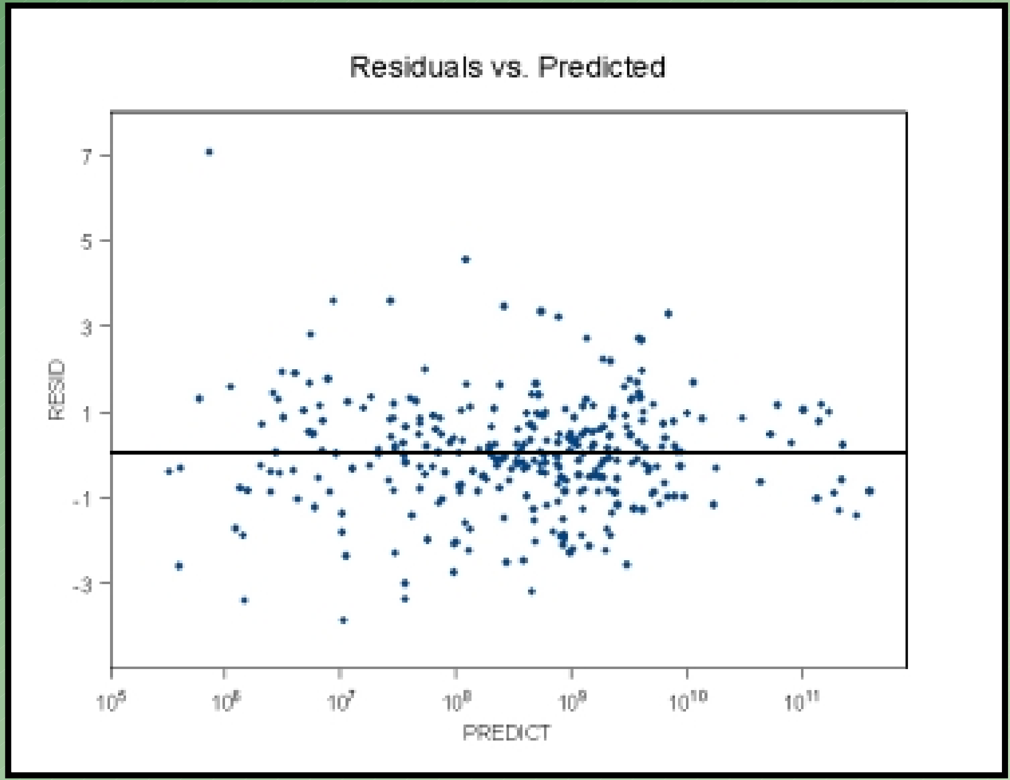


<u>Solution</u>: Weighting of residuals for measurement error (e.g., load estimation error) and model variance—assumes error-variance relations can be defined.

SPARROW Model Calibration Residual Analysis

Fecal Coliform Residual Plot

305 NASQAN sites, records 1978-92

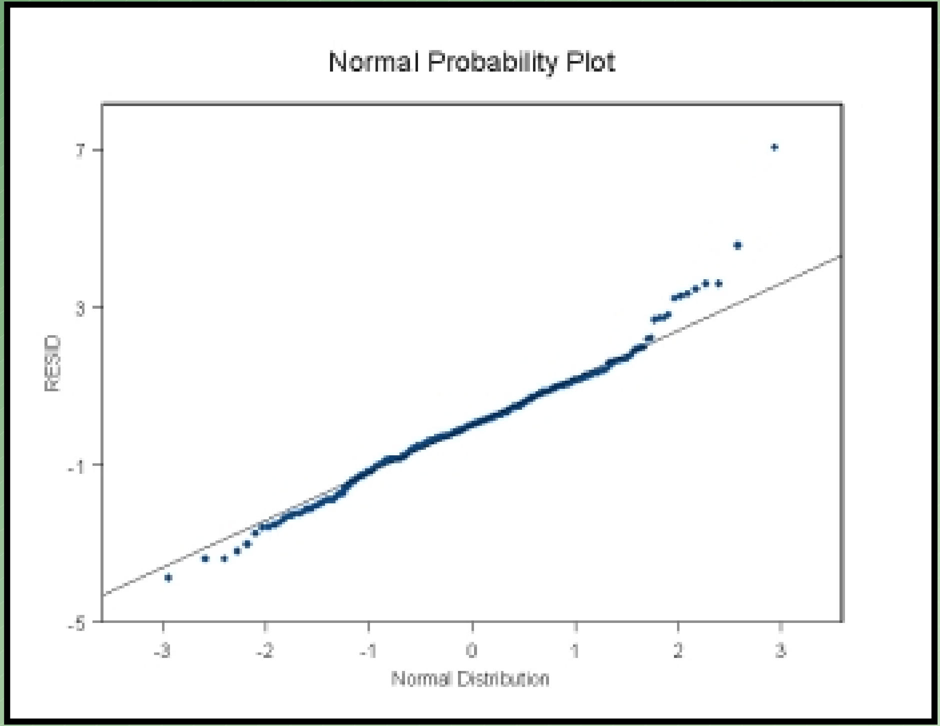


SPARROW Model Calibration Residual Analysis

- Residuals are normally distributed

<u>Fecal Coliform Residual Probability Plot</u>

305 NASQAN sites, records 1978-92



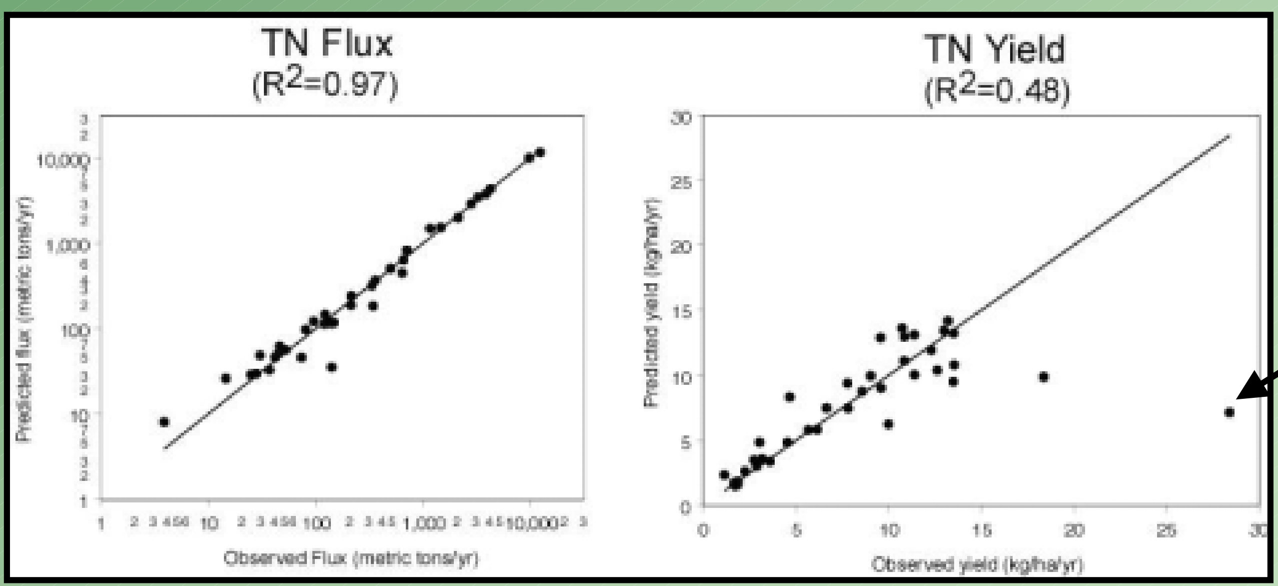
SPARROW Model Calibration Residual Analysis

Outliers

"I like SPARROW's ability to identify inaccuracies in monitoring data…the model is right and the data are wrong."

-Graham McBride, NIWA, NZ

- Caused by data errors or model mis-specification



NZ Waikato TN Model

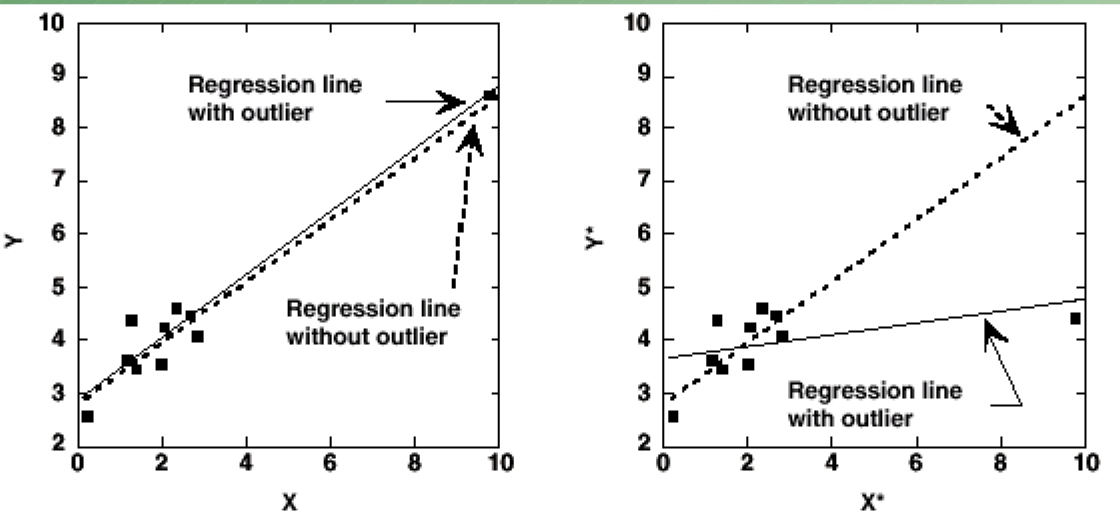
Horticulture prominent in the watershed—source unspecified in the model

SPARROW Model Calibration
Residual Analysis

How Detect Outliers in Multi-Dimensional Space?

• Leverage (hat matrix)—measure of outlier in at least one of the explanatory variables

--High leverage for value > 3p / n,
where p=#parameters; n=# observations



a) outlier with high leverage, but low influence

b) outlier with high leverage and high influence

• Influence statistics (outliers in the response variable):

--Standardized residuals (standard deviation units)

--Cook's D

SPARROW Model Calibration
Model Selection

<u>Significance of model parameters – nested models</u>

- Individual parameters – t statistics (and associated p values)
- Multiple parameters – F test

The test statistic is

$$F = \frac{(SSE_s - SSE_c) / (df_s - df_c)}{(SSE_c / df_c)} \qquad \text{where } (df_s - df_c) = m-k.$$

SPARROW Model Calibration
Model Selection

<u>Significance of model parameters – nested models</u>

- Individual parameters – t statistics (and associated p values)
- Multiple parameters – F test

The test statistic is

$$F = \frac{(SSE_s - SSE_c) / (df_s - df_c)}{(SSE_c / df_c)} \qquad \text{where } (df_s - df_c) = m-k.$$

SSy

SSR (signal)

SSE (noise)

Total sum of squares	=	Treatment sum of squares	+	Error sum of squares
(overall variation)	=	(group means – overall mean)	+	(variation within groups)
$\sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2$	=	$\sum_{j=1}^{k} n_j (\bar{y}_j - \bar{y})^2$	+	$\sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$

USGS
science for a changing world

SPARROW Model Calibration
Model Selection

Overall model fit

- Lowest MSE (Mean Square Error)—declines with increasing # parameters: MSE = SSE / (n-p)
- Highest R-squared—increases with increasing # parameters

$$R^2 = 1 – (SSE / SS_y)$$

where, SS_y = total sum of squares of the regression equation

SSE = sum of square error

- Highest Adjusted R-squared (adjusted for # parameters; relatively insensitive when n>>p):

$$R^2 = 1- [(n-1) SSE] / [(n-p) SS_y]$$

USGS science for a changing world

SPARROW Model Calibration
Model Selection

Overall model fit

- Lowest Mallow's Cp (Cp adjusted for # parameters)
- Lowest PRESS statistic
- Your professional judgement

$$Cp = p + \frac{(n-p) \bullet (s_p^2 - \hat{s}^2)}{\hat{s}^2}$$

where Sp = MSE of the p parameter model

s2 = best estimate of the "true" error (lowest MSE of all models)

SPARROW Model Calibration Summary

Evaluation of model structure and fit

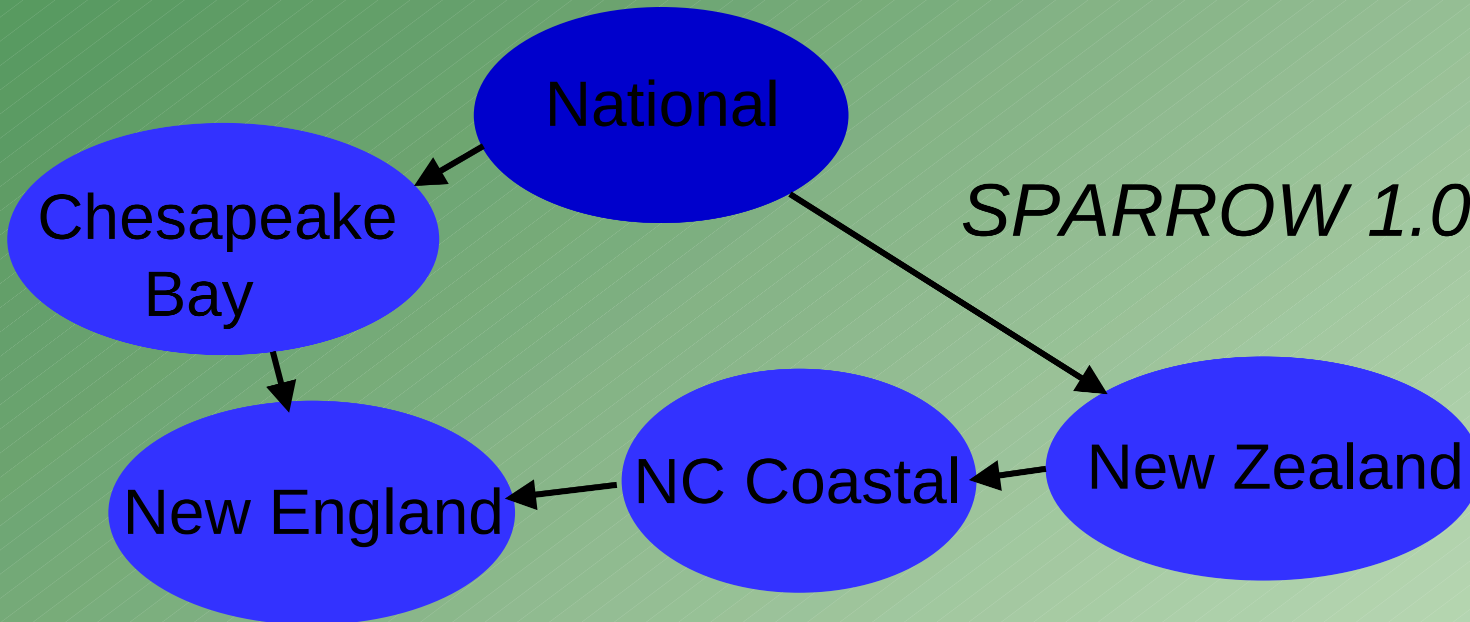
1. Check parameter estimates for statistical and physical significance, correct sign, correlation, and stability

2. Check residual plots for outliers, systematic patterns, homoscedasticity, normality, and inspect mapped residuals

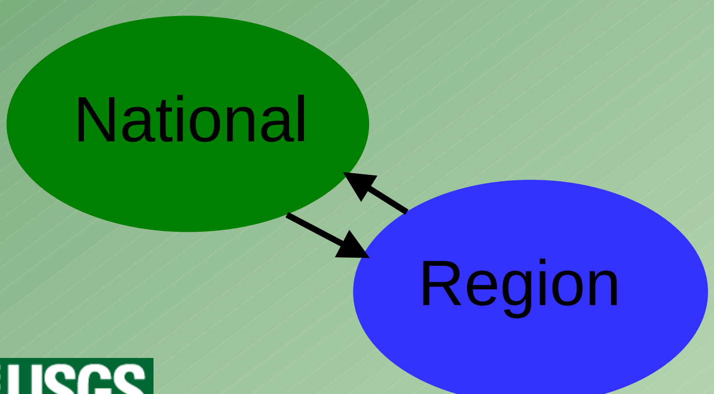
3. Check overall model fit (adjusted R-squared, MSE, Mallow's Cp)

Evolution of SPARROW Calibration Software

SPARROW 1.0



SPARROW 2.0



<u>SAS calibration / prediction software</u>

- New revisions completed and on-going to national code
- Document and support single source of software maintained by the national SPARROW group

<u>GIS methods and software</u> –variety of approaches—will likely continue

SPARROW
SAS Calibration / Prediction Code

- Preprocessing steps:
  - SPARROW explanatory variables identified by reach using GIS
  - Network navigation parameters (hydrologic order of reaches, from- and to-nodes, diversion fraction)
- Assembly of SAS data set (DATA1)
  - Compiles reach-level data from GIS processing in single dataset
- SAS model procedure for calibration and prediction
  - Initial parameter setup (modify variable lists)
  - Data block
    - Final calculations
    - Data screening (e.g., stations)
  - IML procedures (landscape and aquatic decay equations)
  - Output parameters and predictions

USGS
science for a changing world

Fecal Coliform Bacteria
Example Model Building Exercise

Possible sources

Human and animal sources—wastewater, urban runoff, and septic systems, livestock populations (confined feeding operations, unconfined), background for other animals (geese, birds, etc.)

Loss rate for total coliform bacteria:

$k_T = k_B + k_R + k_S$

where,

k_B = base mortality rate (fresh waters) = $0.8 * 1.07^{T-20}$

k_R = solar radiation effect ~ f[light energy (+), depth(-), particulate matter (-)]

k_S = settling loss rate ~ f[particle settling velocity (+), depth(-), fraction of attached bacteria(+)]

USGS
science for a changing world

SPARROW Fecal Coliform Models

Intensive-source model:

Livestock wastes (confined and unconfined), Sewered population, Urban land, Other lands

Land-use (extensive-source) model:

Agricultural lands, Sewered population, Urban land, Other lands

USGS
science for a changing world

SPARROW SAS Calibration / Prediction Code

1. Define response (dependent) and source variables

```
/* Specify all the variable lists. */

/* Dependent variable */
%let depvar = load ;

/* Source variables */
%let srcvar = SEWERPOP RESLAND CONF UNCONF URBAN;

/* Source variable coefficients */
%let bsrcvar = BPOINT BRESLAND BCONF BUNCONF BURBAN;
```

SEWERPOP = sewered population
CONF = confined feeding wastes (kg N)
UNCONF = unconfined feeding wastes (kg N)
URBAN = urban land area (km2)
RESLAND = other land (forest, barren, wetlands, shrub) area (km2)

SPARROW SAS Calibration / Prediction Code

2. Define landscape delivery variables (constrained exponential function)

```
/* Delivery variables */
%let dlvvar = aperm aidrainden  ;

/* Delivery variable coefficients */
%let bdlvvar = bperm bdrainden  ;
```

APERM = Soil permeability (mean adjusted)
AIDRAINDEN = Drainage density (reciprocal mean adjusted)

Adjusted delivery variables (user prompt in future version)

```
/* adjust land-to-water delivery factors by mean */
PROC MEANS DATA=indata ;
 VAR perm idrainden ;
 OUTPUT OUT=mean_ltw MEAN= xperm xdrainden ;
RUN;

data indata; if _n_ = 1 then set mean_ltw; set indata;
  aperm = perm - xperm;
  aidrainden = idrainden - xdrainden;
 run;
```

PERM = Soil permeability
IDRAINDEN = Drainage density (reciprocal)

USGS
science for a changing world

SPARROW SAS Calibration / Prediction Code

3. Define aquatic decay variables

```
/* Decay variables */
%let decvar = rchtot1 rchtot2 rchtot3 ;

/* Decay variable coefficients */
%let bdecvar = brchtot1 brchtot2 brchtot3 ;

/* Reservoir variables */
%let resvar = iresload ;

/* Reservoir variable coefficients */
%let bresvar = bresload ;
```

RCHTOT1 = Reach TOT (days; flow <100cfs)
RCHTOT2 = Reach TOT (days; flow 100 to 500 cfs)
RCHTOT3 = Reach TOT (days; flow >500cfs)

IRESLOAD = Areal hydraulic load (reciprocal; yr/m) for reservoir outlets

Create flow interval variables in data section

```
RCHTOT1 = (meanq <= 100) * (rchtype = 0) * RCHTOT ;
RCHTOT2 = (100 < meanq <= 500) * (rchtype = 0) * RCHTOT ;
RCHTOT3 = (meanq > 500) * (rchtype = 0) * RCHTOT ;

if RHLOAD ^= . and rchtype = 2 then iresload = RHLOAD ;
else iresload = 0 ;
```

MEANQ = mean streamflow (cfs)
RCHTYPE = reach type code
- 0=river reach
- 1=reservoir reach
- 2=reservoir outlet

USGS
science for a changing world

SPARROW SAS Calibration / Prediction Code

4. Define delivery variable design matrix

DLVDSGN Code
0 = delivery not apply to this source
1 = delivery applies to this source

```
/* Specify the delivery design matrix: each row is a different
   source (in the same order as they are listed in the srcvar
   statement); each column is a different delivery variable (in
   the same order as they are listed in the dlvvar statement).
   An element is either a 0 or 1. Element r,c is a 1 if source r
   uses delivery variable c. Otherwise, element r,c is 0.
   A space separates columns and a comma separates rows. */
%let dlvdsgn = 0 0, 1 1, 1 1, 1 1, 1 1;
```

Displayed sequence for DLVDSGN:
0 0 = sewered population for permeability and drainage density

1 1 = residual land area for permeability and drainage density
…etc.

SPARROW SAS Calibration / Prediction Code

5. Select estimation and/or prediction execution mode

```
%let if_estimate = yes ;
%let if_predict = yes ; * Specify if predictions are to be made. ;
%let if_adjust = no  ; * Specify if the load predictions (decayed) are to be adjusted for actual loads at monitoring
stations ;
```

6. Define data set, hydrologic sorting variable, and parameter starting values

```
%let home = d:\model_data\models ;

/* Specify the name of the SAS input data set */
%let indata = DATA2 ;

/* Specify the variable used for ordering the
       reaches in downstream order */
%let seqvar = HYDSEQ ;

/* Specify the parameter list:
      Parameter_name Initial_value Lower_bound:Upper_bound */
%let betailst = bpoint 6991.0 0:.
            bresland 51617.0 0:. bconf 859.0 0:. bunconf 708.0 0:.
            burban 9806138.0 0:.
            bperm 0.301489 0:. bdrainden 0.05563 0:.
            brchtot1 0.67 0:. brchtot2 0.46 0:. brchtot3 0.15 0:.
            bresload 87.0 0:. ;
```

SPARROW SAS Calibration / Prediction Code

7. Define reach and station IDs and network navigation parameters (set once—no need to change)

```
/* Assign a list of column reference vectors */

%let makecol =
  jwaterid = %col(datalst,waterid) %str(;)
  jstatpk  = %col(datalst,statpk) %str(;)
  jfnode   = %col(datalst,fnode) %str(;)
  jtnode   = %col(datalst,tnode) %str(;)
  jfrac    = %col(datalst,frac) %str(;)
  jaiftran = %col(datalst,aiftran) %str(;)
```

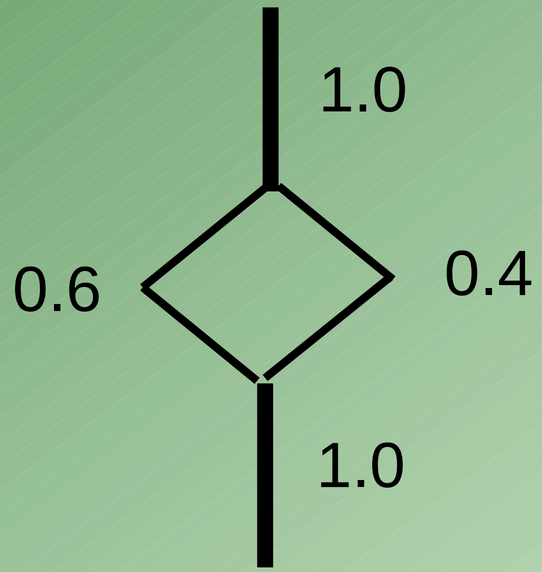
WATERID = watershed / reach ID
STATPK = station ID
FNODE = reach upstream node
TNODE = reach downstream node
FRAC = reach diversion fraction
AIFTRAN = transport flag (1=transport reach)

```
/* Make list of variables to be read from the SAS indata data set and loaded
   into a matrix. Detect and remove duplicates. */

%let addlist = &depvar &srcvar &dlvvar &decvar &resvar &othvar ;
%let datalst = waterid statpk tnode fnode frac aiftran ;
```

SPARROW SAS Calibration / Prediction Code

FRAC – Fraction diversion value



SPARROW SAS Calibration / Prediction Code

7. Define reach and station IDs and network navigation parameters (set once—no need to change or transparent if use same names)

```
/* Assign a list of column reference vectors */

%let makecol =
  jwaterid = %col(datalst,waterid) %str(;)
  jstatpk  = %col(datalst,statpk) %str(;)
  jfnode   = %col(datalst,fnode) %str(;)
  jtnode   = %col(datalst,tnode) %str(;)
  jfrac    = %col(datalst,frac) %str(;)
  jaiftran = %col(datalst,aiftran) %str(;)
```

WATERID = watershed / reach ID
STATPK = station ID
FNODE = reach upstream node
TNODE = reach downstream node
FRAC = reach diversion fraction
AIFTRAN = transport flag (1=transport reach)

```
/* Make list of variables to be read from the SAS indata data set and loaded
   into a matrix. Detect and remove duplicates. */

%let addlist = &depvar &srcvar &dlvvar &decvar &resvar &othvar ;
%let datalst = waterid statpk tnode fnode frac aiftran ;
```

SPARROW SAS Calibration / Prediction Code

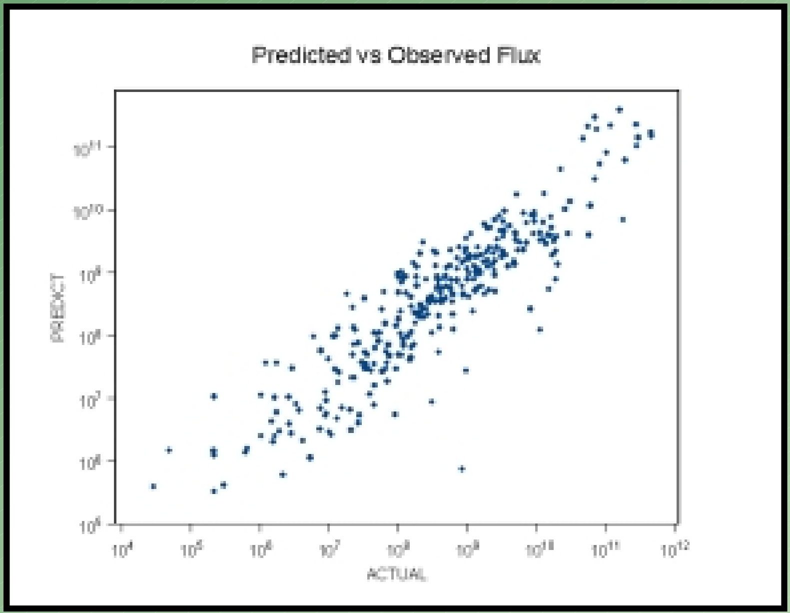
<u>INTENSIVE MODEL RESULTS: Fecal Coliform</u>

305 NASQAN sites, records 1978-92

Sources reflect mean adjusted landscape variables

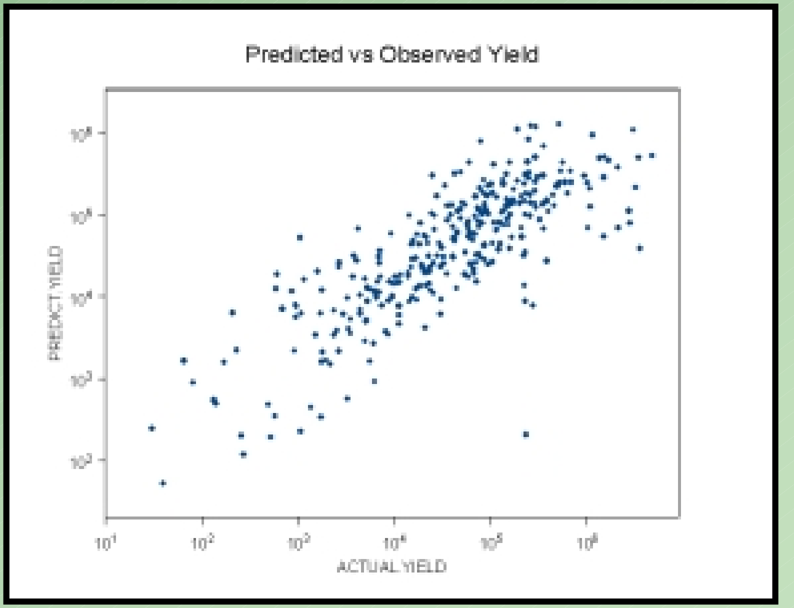
<u>Coliform Flux</u>

(10^{-2} Bcolonies yr^{-1})



<u>Coliform Yield</u>

(10^{-2} Bcolonies km^{-2} yr^{-1})



USGS
science for a changing world

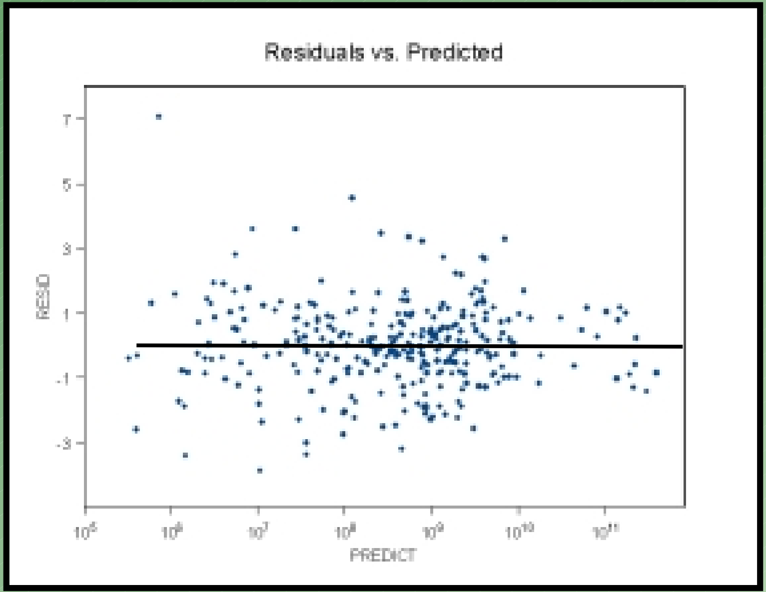
SPARROW SAS Calibration / Prediction Code

INTENSIVE MODEL RESULTS: Fecal Coliform

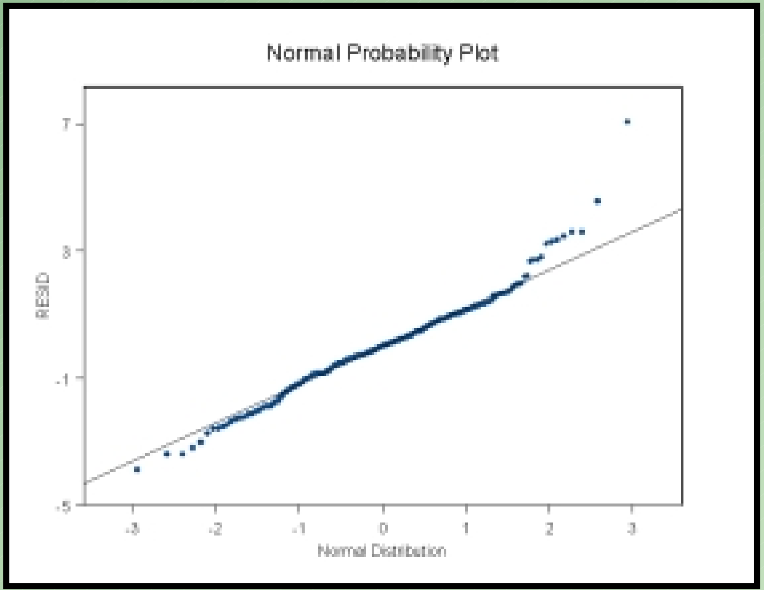
305 NASQAN sites, records 1978-92

Sources reflect mean adjusted landscape variables

Residual Plot



Probability Plot



USGS
science for a changing world

SPARROW SAS Calibration / Prediction Code

INTENSIVE MODEL RESULTS: Fecal Coliform

305 NASQAN sites, records 1978-92

Sources reflect mean adjusted landscape variables

Non-linear Least Squares Results

N Obs	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
305	11	294	549.18903	1.8679899	1.3667443	0.8071341	0.800574

	Parameter	Estimate	Std Err	t Value	Pr > |t|	Coefficient Units
Sources						
Sewered Population	BPOINT	8919.403	2958.4093	3.0149321	0.0027946	8,919/100 = 89.2 Bcol/person/yr
Residual Land	BRESLAND	2045.9055	4741.886	0.4314539	0.6664545	20 Bcol/km2/yr (0.20 Bcol/ha/yr)
Confined wastes	BCONF	324.97938	116.48886	2.7897894	0.0056188	3.25 Bcol/kg N/yr
Unconfined wastes	BUNCONF	177.07559	73.883056	2.3967009	0.0171679	1.77 Bcol/kg N/yr
Urban land	BURBAN	3444100.3	2068444.4	1.6650678	0.0969645	34,441 Bcol/km2/yr (344 Bcol/ha/yr)
Landscape loss						
Soil permeability	BPERM	0.334435	0.0841099	3.976165	0.0000882	h/cm (inverse relation)
Drainage density	BDRAINDEN	0.053385	0.0393925	1.3552088	0.1763913	per km (positive relation)
Aquatic loss						
Stream decay(<100cfs)	BRCHTOT1	0.6485695	0.2210388	2.9341881	0.0036077	per day
Stream decay(100-500cfs)	BRCHTOT2	0.4942857	0.1846064	2.6775113	0.007834	per day
Stream decay(>500cfs)	BRCHTOT3	0.1270183	0.0568133	2.2357144	0.026121	per day
Reservoir decay	BRESLOAD	78.995344	22.727153	3.4758134	0.0005861	m/day

Human fecal coliform intestinal bacteria = 730 Bcol/person/yr

USGS science for a changing world

SPARROW SAS Calibration / Prediction Code

EXTENSIVE (LAND-USE) MODEL RESULTS: Fecal Coliform

305 NASQAN sites, records 1978-92

Sources reflect mean adjusted landscape variables

Non-linear Least Squares Results

N Obs	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
305	10	295	573.11607	1.9427663	1.3938315	0.7987313	0.7925909

	Parameter	Estimate	Std Err	t Value	Pr > |t|	Coefficient Units
Sources						
Sewered Population	BPOINT	9345.4095	3061.8093	3.0522507	0.0024781	9,345/100 = 93.5 Bcol/person/yr
Cultivated Land	BAGRIC	603793.17	185355.63	3.257485	0.0012552	6,038 Bcol/km2/yr (60.4 Bcol/ha/yr)
Residual Land	BRESLAND	6870.043	3741.5327	1.8361574	0.0673406	68.7 Bcol/km2/yr (0.69 Bcol/ha/yr)
Urban land	BURBAN	4251364.3	2457871.4	1.7296936	0.084731	42,514 Bcol/km2/yr (425.1 Bcol/ha/yr)
Landscape loss						
Soil permeability	BPERM	0.2356181	0.0788437	2.9884218	0.00304	h/cm (inverse relation)
Drainage density	BDRAINDEN	0.1459664	0.0786509	1.8558782	0.0644681	per km (positive relation)
Aquatic loss						
Stream decay(<100cfs)	BRCHTOT1	0.6405845	0.2213565	2.8939041	0.0040886	per day
Stream decay(100-500cfs)	BRCHTOT2	0.5339951	0.1996834	2.6742096	0.0079082	per day
Stream decay(>500cfs)	BRCHTOT3	0.126373	0.0601968	2.099331	0.0366373	per day
Reservoir decay	BRESLOAD	70.624909	21.608323	3.2684123	0.0012094	m/day

Human fecal coliform intestinal bacteria = 730 Bcol/person/yr

Cultivated land = NLCD pasture + row crops + fallow land + orchards

USGS science for a changing world

Land-Use Model

Sources	*Bcol yr $^{-1}$*
Sewered population	93.5 person^{-1}
Urban land	425.1 ha^{-1}
Other land	0.69 ha^{-1}
Cultivated land	60.4 ha^{-1}

Land-to-Water Transport:

Soil permeability	(- coef.)
Drainage density	(+ coef.)

In-stream loss (day^{-1})

0.13, 0.53, 0.64

Reservoir loss (m yr^{-1}) 70.6

R-squared 0.79

Reach-accuracy +/- 139%

Intensive-Use Model

Sources:	*Bcol yr $^{-1}$*
Sewered pop.	89.2 person $^{-1}$
Urban land	344.0 ha^{-1}
Other land	0.20 ha^{-1}
Confined waste	3.25 kgN^{-1}
Unconfined waste	1.77 kgN^{-1}

Land-to-Water Transport:

Soil permeability	(- coef.)
Drainage density	(+ coef.)

In-stream loss (day^{-1})

0.13, 0.49, 0.65

Reservoir loss (m yr^{-1}) 79.0

R-squared 0.80

Reach-accuracy +/- 137%

USGS
science for a changing world

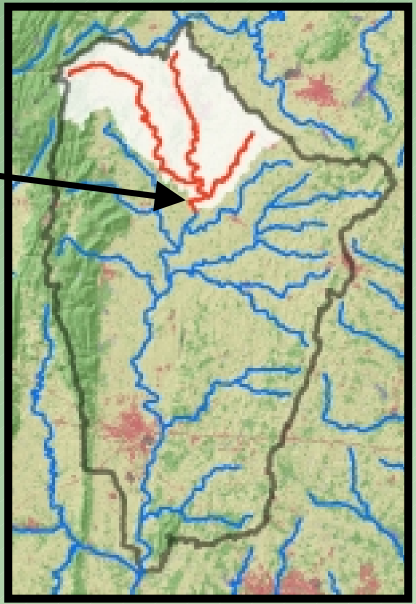
SPARROW SAS Predictions

1. Flux at end of reach from total drainage

PLOAD_TOTAL

PLOAD_"SOURCES"

Units=mass per time



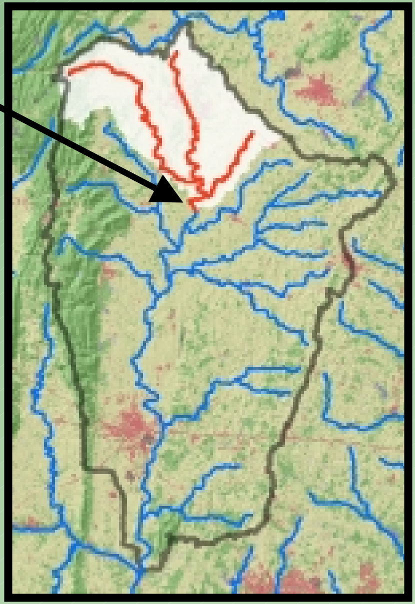
predict_load_lu3c5_update10.sas7bdat

	WATERID	PLOAD_TOTAL	PLOAD_SEWERPOP	PLOAD_RESLAND	PLOAD_CONF	PLOAD_UNCONF	PLOAD_URBAN
1	1	121484650.0	66219466.197	12125206.853	5168037.222	3443474.063	34528465.718
2	2	120829611.6	67246411.492	12328268.739	4890852.6593	3240738.323	33123340.393
3	3	104455083.4	58110091.823	11945150.238	4126711.4436	2697056.243	27576073.743
4	4	34836662.65	11209250.91	12100062.211	955631.90257	410579.0838	10161138.549
5	5	33461.09717	8815.981905	14806.559819	4470.8141532	3177.512838	2190.2284615

USGS
science for a changing world

SPARROW SAS Predictions

2. Flux delivered to end of reach w/o aquatic decay



PLOAD_ND_TOTAL
PLOAD_ND_"SOURCES"

Units=mass per time

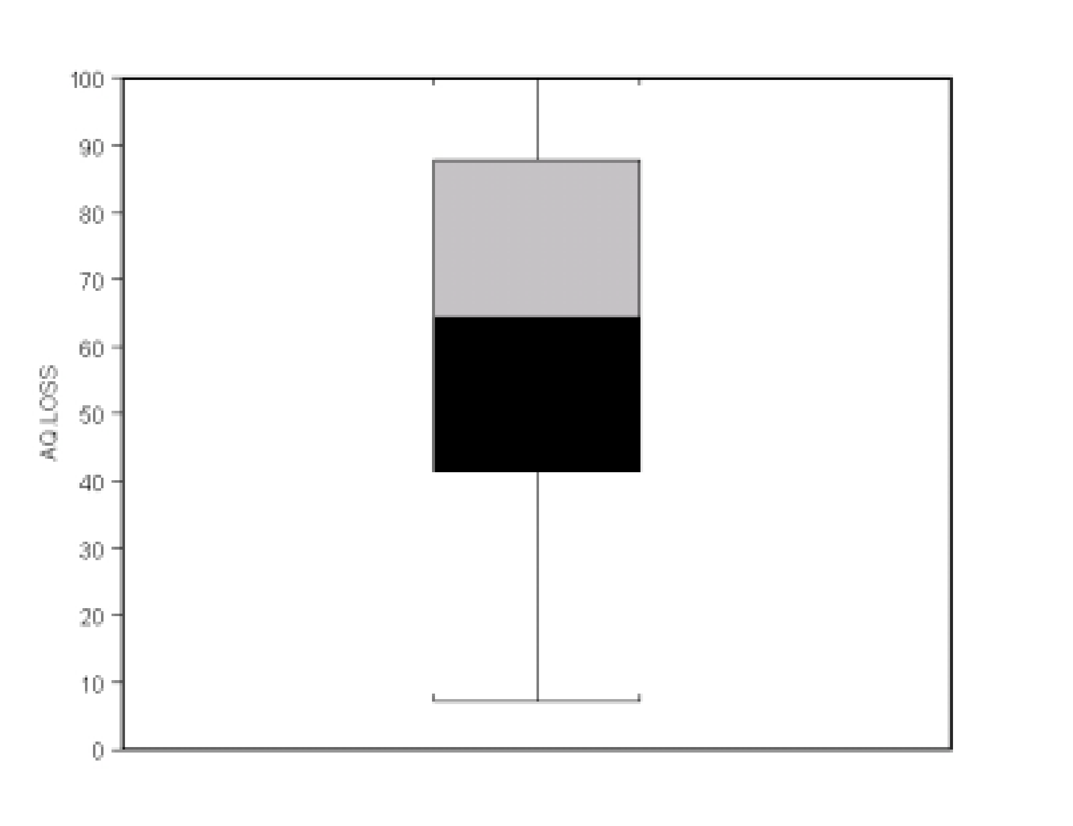
Total mass removed in streams & reservoirs =
PLOAD_ND_TOTAL – PLOAD_TOTAL
% mass removed in stream = (PLOAD_ND_TOTAL – PLOAD_TOTAL) / PLOAD_ND_TOTAL * 100

predict_load_lu3c5_update10.sas7bdat

	PLOAD_ND_TOTAL	PLOAD_ND_SEWERPOP	PLOAD_ND_RESLAND	PLOAD_ND_CONF	PLOAD_ND_UNCONF	PLOAD_ND_URBAN
1	212308549.82	96302794.235	28477245.55	13469327.827	8812938.0104	65246244.192
2	207354644.69	95473289.756	28342708.098	12922726.18	8424455.3553	62191465.303
3	190271689.97	85893850.93	27907222.722	12128509.75	7859987.0468	56482119.517
4	53201191.121	14699176.151	21403030.922	2054432.0414	700102.92756	14344449.079
5	33853.632403	8919.4030041	14980.256942	4523.2616876	3214.7885356	2215.9222342

USGS
science for a changing world

SPARROW Fecal Coliform Model Aquatic Loss (305 NASQAN sites)



SPARROW SAS Predictions

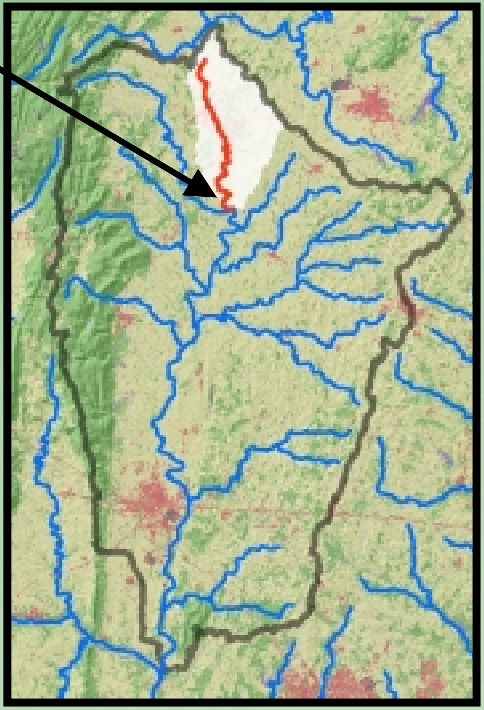
3. Flux delivered to reach from incremental reach catchment

INC_TOTAL
INC_"SOURCES"

Units = mass per time

predict_load_lu3c5_update10.sas7bdat

	INC_TOTAL	INC_SEWERPOP	INC_RESLAND	INC_CONF	INC_UNCONF	INC_URBAN
1	117402.79	53516.418025	1727.7574909	9987.6667	7098.469835	45072.478
2	2690867.4	847343.28539	3413.5396544	18282.879	12994.07299	1808833.6
3	1915108.9	196226.86609	446.97982087	0	0	1718435.1
4	6927415.4	2675820.9012	158071.72354	182251.45	129530.3984	3781740.9
5	33853.632	8919.4030041	14980.256942	4523.2616	3214.788535	2215.9222



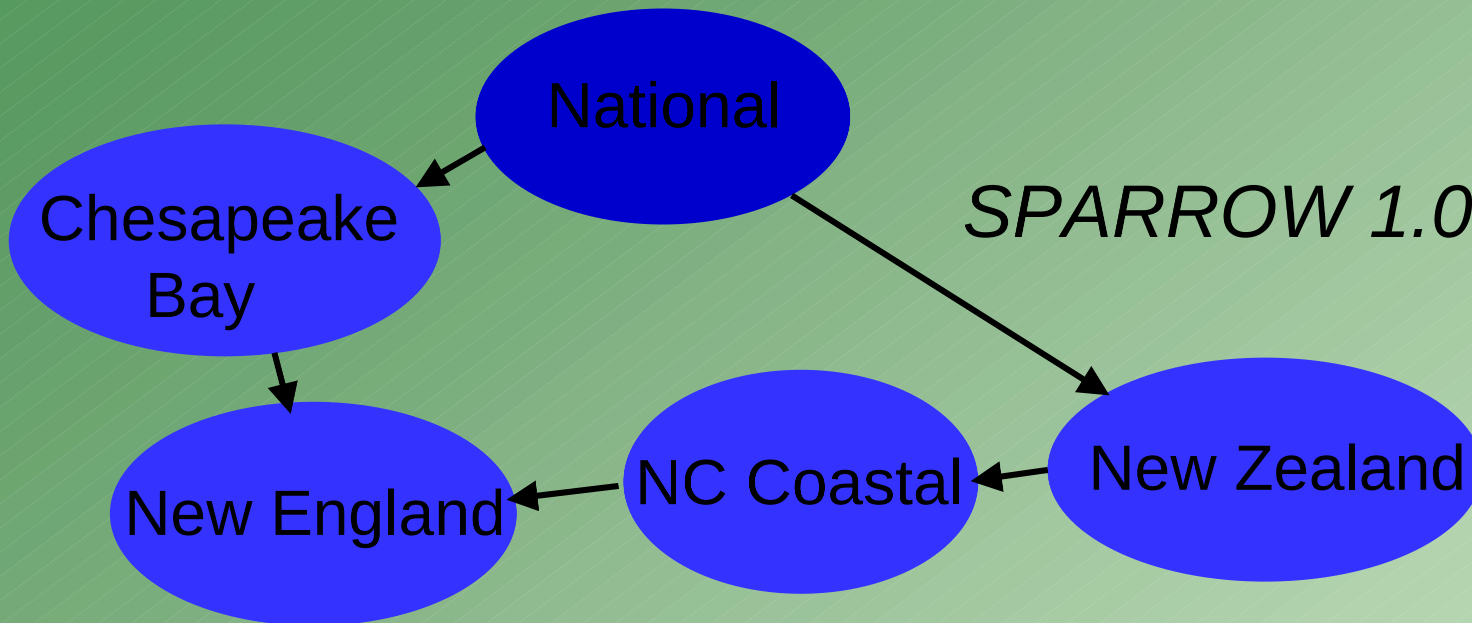
SPARROW SAS Software

Execution run times for new SPARROW code:

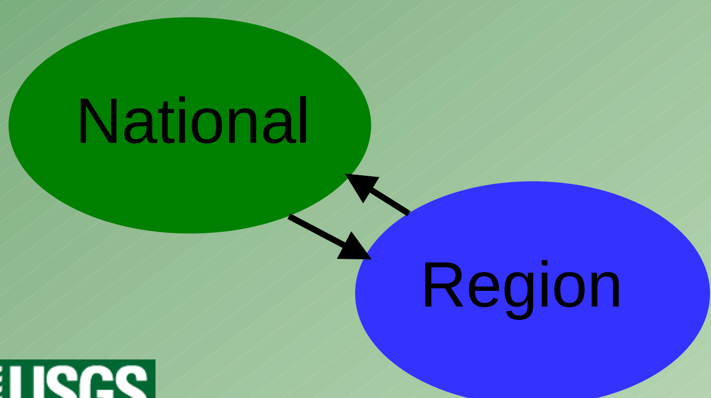
New Zealand Waikato (5,000 reaches) ~ 0.30 minutes

National (65,000 reaches) ~ 10 minutes

Evolution of SPARROW Calibration Software



SPARROW 2.0



<u>SAS calibration / prediction software</u>

- New revisions completed and on-going to national code
- Document and support single source of software maintained by the national SPARROW group

<u>GIS methods and software</u> –variety of approaches—will likely continue

SPARROW SAS Software Enhancements

Remain to be tested (** = needs to be implemented)

1. Delivered flux (incremental, total, sources)
2. Parameter and prediction bootstrapping
3. Prediction confidence intervals
4. Model diagnostics:
   - a. Leverage statistic
   - b. Standardized residuals
   - c. Variance inflation factors (VIFs) **
   - d. Parameter correlation matrix **
   - e. Influence statistics (Cooks D) **
   - f. Mallow's Cp **
5. Weighted observations

   --"measurement error" --load estimation error

   --Hetereoscedasticity **

USGS
science for a changing world